DISSERTATION | DOCTORAL THESIS

Titel | Title

Aspects of volatility modeling: from Gaussian processes to martingales with restricted support

verfasst von | submitted by

Benedict Lukas Bauer BSc. MSc ETH

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 796 605 405
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Mathematik
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Christa Cuchiero Univ.-Prof. Dipl.-Ing. Dr.techn. Stefan Gerhold

Abstract

This thesis examines two distinct lines of research motivated by volatility modeling. The first part presents a characterization result for finite-dimensional self-similar Gaussian Markov processes, supported by our development of a theory on real matrix semigroups. The second part focuses on the existence of martingale measures with restricted support. The theory developed therein is then applied to formulate no-arbitrage conditions in volatility modeling.

Self-similar Gaussian Markov processes
We characterize all multi-dimensional real self-similar Gaussian Markov processes. Three types of covariance matrix functions occur: white-noise type functions, covariances that can be expressed by continuous matrix semigroups, and covariances based on discontinuous solutions of Cauchy's functional equation. Characterizing the latter requires us to develop some results on the representation theory of non-continuous matrix semigroups, satisfying a mild boundedness assumption, without assuming continuity. In addition to the continuous solutions of the semigroup functional equation, we give a description of solutions arising from non-measurable solutions of Cauchy's functional equation. In dimension one, besides white noise, the self-similar Gaussian Markov processes reduce to a two-parameter family of time-changed Brownian motions. This observation simplifies several proofs of non-Markovianity found in the literature.

Martingales with restricted support
In discrete time financial markets, a sufficient condition for the absence of arbitrage is the existence of a martingale pricing measure that satisfies market-imposed constraints. From a model-free perspective, this leads to investigating the existence of martingale couplings with restricted domains. However, conventional convex ordering requirements on marginal distributions are insufficient to guarantee their existence. To address this, we reformulate Strassen's theorem, moving away from the convex envelope approach to one that aligns with the geometry of the prescribed domain. This new formulation gives rise to duality results applicable to robust pricing with fixed marginal constraints. In particular we apply these results to investigate the existence of martingale measures compatible with call options and VIX futures.

Kurzfassung

Diese Dissertation untersucht zwei unterschiedliche Forschungsansätze, die durch die Volatilitätsmodellierung motiviert sind. Der erste Teil präsentiert ein Charakterisierungsergebnis für endlich-dimensionale selbstähnliche gaußsche Markov-Prozesse, unterstützt durch unsere Entwicklung einer Theorie über reelle Matrixhalbgruppen. Der zweite Teil untersucht die Existenz von Martingalmaßen mit eingeschränktem Träger. Die daraus entwickelten Resultate werden dann verwendet, um Bedingungen für Arbitragefreiheit in der Volatilitätsmodellierung zu formulieren.

Selbstähnliche gaußsche Markov-Prozesse
Wir charakterisieren alle mehrdimensionalen, reellwertigen selbstähnlichen gaußschen Markov-Prozesse. Es treten drei Arten von Kovarianzmatrixfunktionen auf: Weißes-Rauschen-Typ-Funktionen, Kovarianzen, die durch stetige Matrixhalbgruppen ausgedrückt werden können, und Kovarianzen, die auf unstetigen Lösungen der Cauchy-Funktionalgleichung basieren. Zur Charakterisierung der letzteren ist es notwendig, einige Resultate zur Darstellungstheorie unstetiger Matrixhalbgruppen zu entwickeln, die eine milde Beschränktheitsannahme erfüllen. Neben den stetigen Lösungen der Halbgruppengleichung beschreiben wir auch Lösungen, die auf nicht-messbaren Lösungen der Cauchy-Funktionalgleichung basieren. Im eindimensionalen Fall reduzieren sich die selbstähnlichen gaußschen Markov-Prozesse, abgesehen vom weißen Rauschen, auf eine zweiparametrige Familie von zeitveränderten Brownschen Bewegungen. Diese Beobachtung vereinfacht mehrere in der Literatur gefundene Beweise für das Fehlen der Markoveigenschaft.

Martingale mit eingeschränktem Träger
In zeitdiskreten Finanzmärkten ist eine hinreichende Bedingung für Arbitragefreiheit die Existenz eines Martingalmaßes, welches vom Markt vorgegebene Randbedingungen erfüllt. Aus einer modellfreien Perspektive führt dies zur Untersuchung der Existenz von Martingalkopplungen mit eingeschränktem Träger. Konventionelle Anforderungen an die konvexe Ordnung von Randverteilungen reichen jedoch nicht aus, um die Existenz solcher Martingalbrücken zu garantieren. Um dies zu lösen, formulieren wir Strassens Theorem neu, mit einem Ansatz, der sich an der Geometrie des vorgeschriebenen Trägers orientiert. Diese Formulierung führt zu Dualitätresultaten, die auf robuste Bepreisung mit festen Randverteilungen anwendbar sind. Insbesondere wenden wir diese Ergebnisse an, um die Existenz von Martingalmaßen zu untersuchen, die mit Kaufoptionen und VIX-Futures kompatibel sind.

Acknowledgements

First and foremost, I would like to thank my supervisors, Stefan Gerhold and Christa Cuchiero, for their ongoing support, guidance, and valuable ideas during times when I was unsure how to proceed. I am also grateful to Alexander Cox and Paul Gassiat for agreeing to review this thesis.

Additionally, I want to extend my thanks to Yuan Chen, Guido Gazzani, Luca Gonzato, Florian Huber, Janka Möller, Francesca Primavera, Stefan Rigger and Viktoria Schildhammer for their companionship. A special thanks to Guido, Janka, Florian, and Francesca for listening and helping with mathematical questions, and to Viktoria for her assistance with administrative issues.

Lastly, I would like to express my gratitude to my parents, Petra and Markus, and to my girlfriend, Sara, for their unwavering support throughout this journey.

Contents

Introduction

This thesis explores two main lines of research. The first part focuses on a characterization result for self-similar Gaussian Markov processes, motivated by the application of some of theses processes in rough volatility modeling. The second part studies martingale measures with restricted support which allows us to formulate a condition, analogous to Strassen's theorem, that ensures the absence of arbitrage in a discrete market model with volatility derivatives.

Self-similar Gaussian Markov processes

In the paper "Volatility is Rough," by Gatheral et al. [39], it has been demonstrated that modeling the driving process of volatility as fractional Brownian motion can improve capturing certain stylized facts.
This in turn has catalyzed the development of rough volatility models, where the volatility is described as the solution to a rough Volterra stochastic differential equation. For an overview on rough volatility models see [12]. One significant challenge with these models is that traditional pricing methods, which rely on Markov theory and its associated partial differential equation framework, are no longer readily applicable. This is due to the fact that fractional Brownian motion, which drives these models, is not inherently a (one-dimensional) Markov process. To overcome this difficulty, Markovian lifts have been introduced, where higher-dimensional auxiliary processes are constructed to recover the Markov property and facilitate the use of classical methods for pricing and hedging options in rough volatility settings [26].
Additionally, recent research has examined the asymptotic behavior of rough volatility, leading to approximation techniques for pricing options [37], [38].
Fractional Brownian motion (fBm) is given by

$$W_t^H = \frac{1}{\Gamma(H+1/2)} \int_{-\infty}^{t} (t-s)^{H-1/2} - (-s)^{H-1/2} \mathbb{1}_{\{s<0\}} dB_t$$

where B is a standard symmetric Brownian motion and $H \in [0,1]$, see for example [58]. Some key properties of fBm are that it is a Gaussian process, meaning that its finite-dimensional distributions are multivariate Gaussian, and it exhibits self-similarity and dependent increments for $H \neq 1/2$. A continuous time process X is self-similar with self-similarity coefficient H if $(X_{at})_{t\geq 0}$ and $(a^H X_t)_{t\geq 0}$ have the same law for any $a > 0$ and we say that it is centered if $\mathbb{E}[X_t] = 0$ for all $t \geq 0$. These properties make fBm particularly suitable for modeling phenomena with memory. For a proof focused solely on the non-Markovianity of fractional Brownian motion, we refer to [45].

A natural extension of the question regarding the Markovian nature of fractional Brownian motion (fBm) is to explore the Markovianity of self-similar centered Gaussian processes which we treat in Chapter 1. An essential tool in the study of Gaussian processes is the associated covariance function, defined as

$$R(s,t) = \mathbb{E}[(X_s - \mu_s)(X_t - \mu_t)],$$

where

$$\mu_t := \mathbb{E}[X_t]$$

is the mean of X at time $t \in \mathbb{R}_{\geq 0}$. The covariance function together with the mean uniquely defines the law of the process. A classical result by Doob [32, Theorem V.8.1] states that a Gaussian process is Markovian if and only if its covariance function satisfies the following identity: for all $0 \leq s \leq t \leq u$ we have

$$R(s,u)R(t,t) = R(s,t)R(t,u). \tag{0.0.1}$$

Therefore, demonstrating that fractional Brownian motion lacks the Markov property reduces to showing that the covariance kernel of fractional Brownian motion given by

$$R_{\text{fBm}}(s,t) = \frac{1}{2}\Big[t^{2H} + s^{2H} - |t-s|^{2H}\Big].$$

does not satisfy (0.0.1). A crucial observation in this context is that if we additionally assume that $R(1,1) \neq 0$, i.e. that the process is not white noise, then the function

$$g(x) = \frac{R(e^{-x},1)}{R(1,1)}$$

must satisfy the functional equation

$$g(x)g(y) = g(x+y). \tag{0.0.2}$$

The most prominent solution to this equation is the exponential function, and it is well known that any measurable solution to this functional equation must indeed be an exponential function. This result primarily stems from the theory on solutions to Cauchy's functional equation. Algebraically, the solution set can be identified with the set of all group homomorphisms from $(\mathbb{R},+)$ into $(\mathbb{R}_+,\times)$ and it heavily relies on the axiom of choice, as any discontinuous solution is also non-measurable. Given that these discontinuous solutions are highly irregular, one can argue that the solutions arising from covariance functions must be continuous. Consequently, g must be an exponential function, which allows for a complete characterization of the covariance functions of self-similar Gaussian Markov processes on $\mathbb{R}$. This analysis yields the family of processes

$$X_t^{H,c} := t^{2H+c}W(t^{-2H-2c})$$

which together with the limiting case $c \to \infty$ exhaust the family of centered self-similar Gaussian Markov processes up to a constant multiple. This characterization is useful as it provides a clear and concise description of the structure of these processes, promoting further analysis and applications as proposed in [33] as a model for anomalous diffusions. Moreover, this result can be applied to obtain a concise proof of the non-Markovian nature of the Riemann-Liouville fractional Brownian motion, a process that has been widely studied in various fields due to its unique properties and applications. Another immediate application is that we also obtain the non Markovianity of the volatility in the rough Bergomi model introduced in [10].

Multi-dimensional case

In Chapter 2 we extend the investigation of the Markov property to multi-dimensional centered self-similar Gaussian processes. A similar line of reasoning as in Chapter 1 can be employed to completely characterize the covariance functions of such processes. However, the analysis becomes considerably more intricate due to the potential appearance of non-measurable covariance functions. In this multi-dimensional setting, the covariance function of a centered Gaussian process X is represented by a unique linear operator $R(s,t)$, defined by the relation.

$$\langle v, R(s,t)u\rangle = \mathbb{E}[\langle X_s, v\rangle\langle X_t, u\rangle]$$

where $s \leq t$ and u, v are vectors in the underlying vector space. This generalizes the scalar case to a framework where the covariance function is now an operator acting on the vector space. Assuming the Markovian property, the covariance operator R can be further reduced to a one-parameter matrix semigroup satisfying (0.0.2). By choosing an appropriate basis of the underlying vector space this facilitates the reduction

$$g(x) = R(e^{-x}, 1)$$

such that g forms a semigroup of matrices, i.e. matrices which satisfy (0.0.2). This matrix semigroup captures the evolution of the long range dependence of the process over time. As in the one dimensional case it has been shown in [51] that any measurable matrix one parameter semigroup must be continuous, however it turns out that we do not necessarily have continuity of g. An essential property of $(g(x))_{x>0}$ stems from the Gaussian and self-similar nature of the underlying process, namely that it must satisfy the boundedness condition

$$||g||_{\text{op}} \leq e^{-Hx} \tag{0.0.1}$$

where H denotes the self-similarity coefficient and $||\cdot||_{\text{op}}$ denotes the operatornorm. We study semigroups satisfying this boundedness condition in Chapter 3. It is well known that any one parameter semigroup on a Banach space which is uniformly continuous can be given as the exponential of its infinitesimal generator. The boundedness condition (0.0.1) allows us to obtain an explicit representation of solutions to the functional equation, analogous to the exponential representation of uniformly continuous semigroups. Specifically, any such solution can by Theorem 3.2.6 be expressed in the form

$$g(x) = Q(x)\exp(Mx)$$

where M is a matrix that plays a role similar to that of the infinitesimal generator in the uniformly continuous case and $Q(x)$ is a semigroup in $SO(n)$, the special orthogonal group on $\mathbb{R}^n$, whose representation depends on discontinuous (and hence non-measurable) solutions of Cauchy's functional equation. In particular $(Q(x))_{x\geq 0}$ are rotation matrices where the angle of rotation $\eta(x)$ solves

$$\eta(x+y) = \eta(x) + \eta(y).$$

This then leads to an explicit representation of the covariance functions of centered Gaussian self-similar Markov processes formulated in Theorem 2.2.15. Due to the measurability challenges, only self-similar Gaussian Markov processes with continuous covariance kernels can be effectively simulated. A Volterra representation of a process X is a representation of the form

$$X_t = \int_0^t K(s,t)dW_s$$

where $K : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}^{d\times d}$ is measurable such that

$$\int_0^T \int_0^t K(s,t)^2 ds dt < \infty$$

for all $T > 0$, and W denotes the standard Brownian motion. A key result in this context is that a Gaussian self-similar Markov process X with self-similarity coefficient H admits a Volterra representation if and only if the following condition holds:

$$\left\| \frac{R(s,t)}{(st)^H} \right\| < 1. \tag{0.0.2}$$

This is shown in Theorem 2.5.5. Condition (0.0.2) arises from the fact that a process possesses a Volterra representation if and only if it is purely non-deterministic, see [24], [49]. In other words, if $\mathcal{F}_t^X$ denotes the filtration generated by the process X at time $t > 0$, then the intersection of these filtrations over all $t > 0$, denoted by $\cap_{t>0}\mathcal{F}_t$, is trivial. This implies that the process has no deterministic component and that its future values cannot be exactly predicted from its past. Conversely if there exists a single pair $s, t \in \mathbb{R}_{>0}$ with $s \neq t$ such that the normalized covariance ratio $R(s,t)/(st)^H$ is an isometry, then it turns out the process X is purely deterministic, see Section 2.5. In this case, for any $0 < s < t$, the sigma-algebra $\mathcal{F}_s^X$ generated by the process up to time s coincides with $\mathcal{F}_t^X$, meaning that the entire future evolution of the process is fully determined by its history up to time s.

In Chapter 4 we present a new proof of the inequality

$$2\Big(1-\cos\frac{\pi}{n+1}\Big)\sum_{k=1}^{n} a_k^2 \le \sum_{k=1}^{n+1}(a_k-a_{k-1})^2 \le 2\Big(1+\cos\frac{\pi}{n+1}\Big)\sum_{k=1}^{n} a_k^2. \tag{0.0.3}$$

for real $a_1,\dots,a_n$ and $a_0 = a_{n+1} = 0$, originally due to [36] and [56]. We relate this inequality to the dissipativity (see 4.2.1) of a Jordan block and further provide an extension via the Lumer-Phillips theorem 4.3.1. Additionally, we obtain a slight generalization of the Lumer-Phillips theorem in finite dimensions, which is then applied in the construction of the Volterra representation in Theorem 2.4.3.

Martingales with restricted support

A fundamental result in martingale theory is Strassen's theorem, which provides a criterion for the existence of a martingale coupling between two given probability distributions. Specifically, if μ and ν are two probability measures on $\mathbb{R}^d$, then a martingale coupling exists between these measures if and only if, for all non-negative convex functions $f : \mathbb{R}^d \to \mathbb{R}$, the following inequality holds:

$$\mathbb{E}^\mu[f] \le \mathbb{E}^\nu[f].$$

Now, consider the problem of determining when there exists an n-step martingale with a given initial distribution μ and a final marginal distribution ν. If the conditions of Strassen's theorem are satisfied, such a martingale can indeed be constructed. The construction process involves selecting a one-step martingale coupling that satisfies the given distributions, followed by employing the pairwise identity coupling for the remaining steps. This method constructs a Markov martingale with marginals $\mu, \nu, \dots, \nu$. However, the situation becomes more intricate when we impose restrictions on the paths that the martingale can take. Specifically, if the domain of permissible paths is restricted, the existence of a martingale coupling in the multi-step problem requires more careful consideration, as the identity coupling might no longer be feasible. In this context, restricting the path space means that there exists a set of allowed paths P such that we are interested only in martingale measures $\mathbb{Q}$ satisfying $\mathbb{Q}(P) = 1$. To formulate a Strassen-type theorem for the existence of such martingale measures under path restrictions, we turn to a variation of Strassen's original theorem that involves the concept of the convex envelope of a function. For any function $f : \mathbb{R}^d \to \mathbb{R}$ let $\mathcal{L}(f)$ denote the largest lower semi-continuous convex function dominated by f, see Definition 5.A.5. Strassen's theorem can then be reinterpreted as stating that the following condition is equivalent to the existence of a martingale coupling:

$$\mathbb{E}^\mu[\mathcal{L}f] \le \mathbb{E}^\nu[f]$$

for all real-valued non negative continuous functions f. Our contribution is the construction of an operator

$$S_P : (f : \mathbb{R}^d \to \mathbb{R}) \to (S_P f : \mathbb{R}^d \to \mathbb{R})$$

depending on P such that there exists a martingale measure along P with initial distribution μ and terminal distribution ν if and only if

$$\mathbb{E}^\mu[S_P f] \leq \mathbb{E}^\nu[f] \tag{0.0.1}$$

holds for a suitable class of test functions. In the two step case the operator S_P can be obtained from weak optimal transport (for an overview see [9]). Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ denote the extended reals and denote by $\mathcal{P}^1(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$ with finite first moment. Given a closed set $P \subset \mathbb{R}^d \times \mathbb{R}^d$ we set $P_x = \{y : \ (x, y) \in P\}$. Defining the cost function $c : \mathbb{R}^d \times \mathcal{P}^1(\mathbb{R}^d) \to \overline{\mathbb{R}}$ by

$$c(x, \rho) := \begin{cases} 0 & \text{if } \operatorname{supp}(\rho) \subset P_x \text{ and } \operatorname{barycenter}(\rho) = x \\ \infty & \text{else.} \end{cases}$$

The optimization problem,

$$\inf_{\nu = \pi_x * \mu} \int c(x, \pi_x) \mu(dx),$$

where $\pi_x * \mu$ denotes the convolution, is 0 if and only if there exists a martingale measure $\mathbb{Q}$ with marginal μ, ν and $\mathbb{Q}(P) = 1$. The weak optimal transport duality in [8] then yields that there exists a desired martingale coupling if and only if

$$\sup_{g \in L^1(\mu), f} \mathbb{E}^\mu[g] - \mathbb{E}^\nu[f] = 0 \tag{0.0.2}$$

where the supremum is taken over all linearly bounded continuous functions f and functions $g \in L^1(\mu)$ satisfying

$$g(x) - \mathbb{E}^\rho[f] \leq c(x, \rho) \tag{0.0.3}$$

for all $(x, \rho) \in \mathbb{R}^d \times \mathcal{P}^1(\mathbb{R}^d)$. This is equivalent to

$$g(x) \leq \mathbb{E}^\rho[f] \ \text{ for } \operatorname{supp}(\rho) \subset P_x \text{ and } \operatorname{barycenter}(\rho) = x.$$

Since this needs to hold for all such ρ we must have

$$g(x) \leq \inf_{\substack{\rho,\ \operatorname{supp}(\rho) \subset P_x \\ \operatorname{barycenter}(\rho) = x}} \mathbb{E}^\rho[f].$$

Setting

$$S_P f(x) := \inf_{\substack{\rho,\ \operatorname{supp}(\rho) \subset P_x \\ \operatorname{barycenter}(\rho) = x}} \mathbb{E}^\rho[f]$$

and assuming the integrability and measurability of $S_P f$ it follows from (0.0.2) that if for all linearly bounded functions f we have the inequality

$$\mathbb{E}^\mu[S_P f] \leq \mathbb{E}^\nu[f]$$

then there exists a martingale measure with marginal μ and ν which is supported on P. It turns out that the above inequality condition is also necessary and hence this result mirrors the structure of Strassen's theorem. There is also a dual representation for the operator S_P, namely

$$S_P f(x) := \sup\{g(x) |\ g : \mathbb{R}^d \to \mathbb{R} \text{ convex and } g(y) \leq f(y) \text{ for } y \in P_x\}.$$

In the multi step setting it becomes more difficult to employ the theory of weak optimal transport. Theorem 5.2.9 shows that condition (0.0.1) is also sufficient in the multistep case for S_P introduced in Definition 5.2.4. Additionally we obtain duality results on the marginals of such martingale couplings leading to robust pricing bounds even without path restrictions, see Theorem 5.2.9, Theorem 5.2.14 and Remark 5.2.15.

Martingale couplings compatible with call options and VIX futures

The operator S_P can be leveraged to derive a Strassen-type existence result for martingale measures that are compatible with both call option prices on the SPX and the VIX [1]. The reason to focus on call options stems from the classical fact that their prices completely determine the marginals of the pricing measure (provided that they are available for all strikes). From a theoretical viewpoint the VIX is given as

$$\mathrm{VIX}_t^2 = \frac{2}{30}\mathbb{E}^{\mathbb{P}}\left[\log\left(\frac{S_t}{S_{t+30}}\right)\middle| \mathcal{F}_t\right]$$

where $\mathbb{P}$ denotes the pricing measure, S_t, S_{t+30} the prices of the SPX at time t and $t+30$, and $\mathcal{F}_t$ denotes the available information up to time t. It is a measure for the markets expectation of volatility for the coming 30 days.
A longstanding problem, which has seen significant recent progress (see [40], [25], [60]), is the development of market models that calibrate to both SPX and VIX smiles. Assuming that there are no arbitrage opportunities in the SPX and VIX, by [42] any market model calibrated to the SPX and VIX yields for any t a discrete measure μ on $(S_t, S_{t+30}, \mathrm{VIX}_t^2)$ such that

$$\mathbb{E}^\mu[S_{t+30}|S_t, \mathrm{VIX}_t] = S_t \quad \text{and} \quad \mathrm{VIX}_t^2 = \frac{2}{30}\mathbb{E}^\mu\left[\log\left(\frac{S_t}{S_{t+30}}\right)\middle| S_t, \mathrm{VIX}_t\right].$$

Conversely the existence of such a measure is equivalent to the absence of arbitrage with respect to semi-static trading strategies. In [42] this resulting two-step model is studied in more detail. In discrete markets, the absence of arbitrage has been extensively studied in the context of robust pricing; see for example [2] and [14]. For the continuous case, we refer to [3], [13], [22] and [31]. In this thesis we present a condition akin to Strassen's theorem guaranteeing the existence of such a pricing measure. Denote by μ_1, μ_2 and μ_3 the pricing measures of S_t, VIX_t^2 and S_{t+30} respectively. In Theorem 5.3.7, by setting

$$P := \{((s_1, c), (s_2, 2/30 \log(s_1/s_2)))|\ s_1, s_2 \in \mathbb{R}_{>0}, c \in \mathbb{R}\}$$

and applying the operator S_P adapted to the current setup, we find that an arbitrage-free pricing measure exists if and only if for all non negative continuous functions f and bounded continuous functions ϕ_1, ϕ_2 satisfying $\phi_1 \oplus \phi_2 \leq S_P f$ we have

$$\mathbb{E}^{\mu_1}[\phi_1] + \mathbb{E}^{\mu_2}[\phi_2] \leq \mathbb{E}^{\mu_3}[f].$$

Structure of the thesis

The first part is dedicated to our work on self-similar Gaussian Markov processes. In Chapter 1, we focus on the characterization of these processes in the one-dimensional case, while Chapter 2 extends this analysis to the multi-dimensional case. Chapter 3 presents the proof of the representation result for bounded semigroups, which is central to the proof of the main result in Chapter 2. In Chapter 4, we prove (0.0.3) and explore its applications to self-similar Gaussian Markov processes. The second part of the thesis, consisting solely of Chapter 5, focuses on martingales with restricted support and their financial applications.

Bibliographic information on papers contained in this thesis

Chapters 2, 3, 4, and 5 are based on the following papers below. While these chapters closely follow the content and structure of the original works, minor modifications have been made to better integrate them into the overall thesis. Additionally, Section 2.5 contains new material.

Paper 1. (Chapter 2)

A characterization of real semigroups
Bauer Benedict, Gerhold Stefan.
Research in Mathematics.
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.
Online versions available at: https://arxiv.org/abs/2305.15522

Paper 2. (Chapter 3)

Self-similar Gaussian Markov processes
Bauer Benedict, Gerhold Stefan.
not published elsewhere yet
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.

Paper 3. (Chapter 4)

The Fan–Taussky–Todd inequalities and the Lumer–Phillips theorem
Bauer Benedict, Gerhold Stefan.
Journal of Inequalities and Special Functions.
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.
Online versions available at: https://arxiv.org/abs/2305.14015

Paper 4. (Chapter 5)

Martingales with restricted support
Bauer Benedict, Cuchiero Christa.
not published elsewhere yet
Contribution of the author of this dissertation:
The paper was worked out in close collaboration.

Part I.

Self-similar Gaussian Markov processes

1. Self-similar Gaussian Markov processes in one dimension

1.1. Introduction

Fractional Brownian motion with Hurst parameter H is among the best-known examples of H-self-similar Gaussian stochastic processes. It is not Markovian for $\frac{1}{2} \neq H \in (0,1)$, which can be proven using the functional equation that the covariance function of any Gaussian Markov process must satisfy (see [46] and [58, Theorem 2.3]). Many authors have considered self-similar Gaussian processes that are variants and extensions of fractional Brownian motion [18, 65, 66, 67, 78], and for some, the natural question of Markovianity has been studied. In dimension one, it is known that all Gaussian Markov processes are time-changed Brownian motions [6, p. 4]. Our contribution in dimension one is that we explicitly determine this time change for self-similar processes; it depends on the self-similarity parameter and an additional parameter (Theorem 2.6.1). It is usually a very simple matter to compare the resulting explicit covariance function with a given covariance function, yielding easy non-Markovianity proofs for many self-similar Gaussian processes that have been considered in the literature. Moreover the two-parameter family of Gaussian processes we just mentioned has been analyzed in detail in [33], where it is proposed as a model for anomalous diffusion.

1.2. Characterization in the one dimensional case

For $H > 0$ and $c \leq -H$, the process

$$X_t^{H,c} := t^{2H+c} W(t^{-2H-2c}), \quad t \geq 0, \tag{1.2.1}$$

where $W(\cdot)$ is a Brownian motion, is clearly a Gaussian Markov process. By Brownian scaling, it is easy to check that $X^{H,c}$ is H-self-similar. The main result of the present chapter is that this family of processes, augmented by the limiting case $c \to -\infty$, is not just a subset but *equal to* the class of self-similar centered Gaussian Markov processes (up to multiplication by constants). For the notion of self-similarity, we use the same definition as [34] (pp. 1–3).

Definition 1.2.1. *A stochastic process $(X_t)_{t\geq 0}$ is H-self-similar with exponent $H > 0$, if it is stochastically continuous at zero, and for any $a > 0$ the process $(X_{at})_{t\geq 0}$ has the same law as $(a^H X_t)_{t\geq 0}$.*

Self-similarity implies that $X_0 = 0$ (see p. 2 in [34]). In particular, the covariance function of a self-similar process satisfies $R(s,t) = 0$ for $s \wedge t = 0$.

Definition 1.2.2. *A symmetric function $R : [0,\infty)^2 \to \mathbb{R}$ is positive definite if*

$$\sum_{k,l=1}^{d} a_k a_l R(t_k, t_l) \geq 0 \tag{1.2.2}$$

for all $d \geq 1$, $t_1, \ldots, t_d \geq 0$ and $a_1, \ldots, a_d \in \mathbb{R}$.

It is well known that positive definite functions are exactly the functions that occur as covariance functions of centered Gaussian processes $(X_t)_{t\geq 0}$. The covariance function of $X^{H,c}$ is

$$R_{H,c}(s,t) := \begin{cases} (s \vee t)^{2H+c}(s \wedge t)^{-c} & s \wedge t > 0, \\ 0 & s \wedge t = 0. \end{cases} \tag{1.2.3}$$

The pointwise limit

$$R_{H,-\infty}(s,t) := \lim_{c \to -\infty} R_{H,c}(s,t) = \begin{cases} t^{2H} & s = t, \\ 0 & s \neq t, \end{cases} \tag{1.2.4}$$

is positive definite as well, and defines a centered Gaussian process $X^{H,-\infty}$, which is obviously also self-similar and Markov. Note that putting $H = \frac{1}{2}$ and $c = -1$ in (1.2.3) yields the covariance function $s \wedge t$ of Brownian motion, the prime example of a self-similar Gaussian Markov process. We can now state our main result.

Theorem 1.2.3. *Let $H > 0$ and $X = (X_t)_{t\geq 0}$ be a one-dimensional, centered H-self-similar Gaussian Markov process. Then the covariance function of X is of the form*

$$R(s,t) = R(1,1) R_{H,c}(s,t), \quad s,t \geq 0, \tag{1.2.5}$$

where $c \in [-\infty, -H]$ and $R(1,1) \geq 0$. Thus, X equals $R(1,1)^{1/2} X^{H,c}$ in distribution.

Since comparing a given covariance function with the simple explicit function $R_{H,c}$ is typically very easy, it seems that Theorem 1.2.3 essentially settles the problem of deciding whether a self-similar Gaussian process defined by an explicit covariance function is Markovian.

It is well known that a centered real Gaussian process $(X_t)_{t\geq 0}$ is Markov if and only if its covariance function $R(s,t) = \mathbb{E}[X_s X_t]$ satisfies

$$R(s,u)R(t,t) = R(s,t)R(t,u), \quad 0 \leq s \leq t \leq u. \tag{1.2.6}$$

This is due to Doob; we refer to [55] for details and references. To prove Theorem 1.2.3, we use some facts about additive functions. For a proof of the following classical result, and further references, we refer to [16] (Theorem 1.1.7).

Theorem 1.2.4 (Ostrowski 1929)**.** *Suppose that $f:\mathbb{R}\to\mathbb{R}$ satisfies Cauchy's functional equation*

$$f(x+y)=f(x)+f(y), \quad x,y\in\mathbb{R}.$$

Then either $f(x)=cx$ for some constant c, or else f is unbounded above and below on any set of positive measure.

We will also need the following continuation result.

Lemma 1.2.5. *Suppose that $f:[0,\infty)\to\mathbb{R}$ satisfies Cauchy's functional equation*

$$f(x+y)=f(x)+f(y), \quad x,y\ge 0.$$

Then f has an extension to $\mathbb{R}$ that satisfies the equation, too.

Proof. We define f for $x<0$ by

$$f(x):=f(x+y)-f(y),$$

where $y\ge|x|$ is arbitrary. This is well-defined, because for $y_1>y_2\ge|x|$ we have

$$\begin{aligned} f(x+y_2)-f(y_2)&=f(x+y_2)+f(y_1-y_2)-\big(f(y_2)+f(y_1-y_2)\big)\\ &=f(x+y_1)-f(y_1). \end{aligned}$$

By definition, this extension solves the equation for $x<0$ and $y\ge|x|$. Now let $x<0$ and $y<|x|$. Fix z_1 and z_2 with $z_1\ge|x|$ and $z_2\ge|y|$. Then we compute

$$\begin{aligned} f(x+y)&=f(x+y+z_1+z_2)-f(z_1+z_2)\\ &=f(x+z_1)+f(y+z_2)-\big(f(z_1)+f(z_2)\big)\\ &=f(x+z_1)-f(z_1)+f(y+z_2)-f(z_2)=f(x)+f(y). \end{aligned}$$ □

We can now prove Theorem 1.2.3. The proof is an extension of the proof that fractional Brownian motion is not a Markov process, which is found in [46] and [58] (Theorem 2.3).

Proof of Theorem 1.2.3. By self-similarity and symmetry of $R(\cdot,\cdot)$, it suffices to determine $R(\cdot,1)$ on $(0,1]$. Using (1.2.6) (with $u=1$) and self-similarity, we obtain

$$R(s,1)R(1,1)=R(s/t,1)R(t,1), \quad 0\le s<t\le 1. \tag{1.2.7}$$

Putting $s=t^2$, it follows that

$$R(t^2,1)R(1,1)=R(t,1)^2, \quad 0\le t\le 1. \tag{1.2.8}$$

If $R(1,1)=0$, then this equation implies that $R(\cdot,1)$ vanishes, hence X is identically zero. If $R(1,1)\neq 0$, then (1.2.8) implies that $R(\cdot,1)$ is either non-negative or non-positive on $[0,1]$. In the latter case, by putting $a_{kl}=1$ in (1.2.2), we see that $R(\cdot,1)\equiv 0$. We conclude that $R(\cdot,1)$ is non-negative, and may assume from now on that $R(1,1) > 0$. Define

$$g(x) := \frac{R(e^{-x},1)}{R(1,1)}, \quad x \geq 0.$$

Writing $s = e^{-x-y}$ and $t = e^{-y}$, we see from (1.2.7) that g solves the functional equation

$$g(x+y) = g(x)g(y), \quad x, y \geq 0. \tag{1.2.9}$$

In particular,

$$g(z) = g(z/2)^2, \quad z \geq 0. \tag{1.2.10}$$

Clearly, $g(0) = 1 > 0$. By (1.2.9), if x_0 is a zero of g, then g vanishes on $[x_0, \infty)$. Suppose that there is no interval $[0, \varepsilon)$ with $\varepsilon > 0$ on which g is positive. Then, $g(x) = 0$ for $x > 0$, and thus $R = R(1,1)R_{H,-\infty}$, the white-noise-type covariance function defined in (1.2.4). If, on the other hand, there is such an interval $[0, \varepsilon)$, then (1.2.10) yields

$$g(z) = g(z/2^k)^{2^k}, \quad k \in \mathbb{N},$$

and by taking k sufficiently large we get $g(z) > 0$ for any $z \geq 0$. Therefore, $\varphi := \log g$ is well-defined on $[0, \infty)$, and satisfies Cauchy's functional equation

$$\varphi(x+y) = \varphi(x) + \varphi(y), \quad x \geq y \geq 0.$$

By symmetry of the equation, it clearly holds for $0 \leq x < y$ as well. By Lemma 1.2.5, we can extend the solution φ to the whole real line, which makes Theorem 1.2.4 applicable. Suppose for contradiction that φ is unbounded above and below on any set of positive measure. This would imply

$$\sup \big\{ R(t_1/t_2, 1) : 1 \leq t_1 < t_2 \leq 2 \big\} = \infty.$$

We can thus pick $1 \leq t_1 < t_2 \leq 2$ such that

$$R(t_1/t_2, 1) > 2^{2H} R(1,1).$$

Then positive definiteness is violated for $a_1 = 1, a_2 = -1$:

$$\begin{aligned} a_1^2 R(t_1,t_1) + 2a_1a_2R(t_1,t_2) + a_2^2 R(t_2,t_2) &= \\ t_1^{2H}R(1,1) - 2t_2^{2H}R(t_1/t_2,1) + t_2^{2H}R(1,1) & \\ < 2^{2H}R(1,1) - 2^{2H+1}R(1,1) + 2^{2H}R(1,1) &\leq 0. \end{aligned}$$

We deduce that the other possibility stated in Theorem 1.2.4 must hold, i.e. that there is a constant c such that

$$\varphi(x) = cx, \quad x \in \mathbb{R}.$$

By the definition of φ, we obtain

$$R(t,1) = R(1,1)t^{-c}, \quad 0 < t \le 1,$$

and, by self-similarity,

$$\begin{aligned} R(s,t) &= (s \vee t)^{2H} R\Big(\frac{s \wedge t}{s \vee t}, 1\Big) \\ &= R(1,1)(s \vee t)^{2H+c}(s \wedge t)^{-c}, \quad s,t > 0. \end{aligned}$$

For $c > -H$, it is straightforward to check that $(s \vee t)^{2H+c}(s \wedge t)^{-c}$ does not satisfy the Cauchy-Schwarz inequality, hence it cannot be a covariance function. We conclude that $c \le -H$, and that $R = R_{H,c}$. □

1.A. Three additional proofs

In this section we give a variant of the proof of Theorem 1.2.3, show by a non-probabilistic argument that $R_{H,c}$ is positive definite, and discuss the Volterra representation of $X^{H,c}$.

A variant of the proof of Theorem 1.2.3. After settling the case $c = -\infty$, we may assume that R is positive, except for $s \wedge t = 0$. There is an alternative argument to finish the proof in this case. Gaussian Markov covariance functions that have no zeros are always of the form

$$R(s,t) = G(s \wedge t)F(s \vee t); \tag{1.A.1}$$

see [19] and p. 11 in [52]. This result is not immediately applicable, because our covariance function vanishes on the axes. But it suffices to define $H(0) = 0$ in the proof on the first page of [19] (H is our F) to see that (1.A.1) holds, with $G(0) = F(0) = 0$, $G, H > 0$ on $(0,\infty)$, and G/H non-decreasing on $(0,\infty)$. Self-similarity yields

$$G(xs)F(xt) = x^{2H}G(s)F(t), \quad 0 \le s \le t,\ x \ge 0. \tag{1.A.2}$$

Define $A(y) := \log G(e^y)$ and $B(y) := \log F(e^y)$, $y \in \mathbb{R}$. Then (1.A.2) implies

$$A(y+u) + B(y+v) = 2Hy + A(u) + B(v), \quad u \le v,\ y \in \mathbb{R}. \tag{1.A.3}$$

The function $A - B$ is non-decreasing and hence almost everywhere differentiable. Setting $v = y + u$ we obtain

$$A(y+u) - B(y+u) = 2Hy + A(u) - B(2y+u), \quad u \in \mathbb{R},\ y \ge 0,$$

which shows that B, and hence also A, is almost everywhere differentiable. For $n \in \mathbb{N}$ choose $v_n > n$ such that B is differentiable at each v_n. From (1.A.3) we have

$$A(y+u) - A(u) = 2Hy - (B(y+v_n) - B(v_n))$$

for $u < v_n$. This implies

$$A'(u) = 2H - B'(v_n),$$

and hence A has constant derivative for $u < v_n$. As this holds for all $n \in \mathbb{N}$, the function A has constant derivative on the whole real line. Thus, A is linear, and hence G is a power function on $(0, \infty)$, say $G(x) = G(1)x^{-c}$ with $c \in \mathbb{R}$. Since

$$G(x)F(x) = x^{2H}F(1)G(1),$$

$F(x) = F(1)x^{2H+c}$ is a power function on $(0, \infty)$ as well, and we obtain $R = R(1,1)R_{H,c}$. $\square$

The following theorem is a special case of the main theorem in [53]. For background on lattices and Möbius functions (which we do not require here), we refer to chapter 3 in [70].

Theorem 1.A.1 (Lindström 1969). *Let $\{x_1, \dots, x_d\}$ be a set of real numbers in increasing order, $x_1 \le x_2 \le \dots \le x_d$. For functions f_i defined on $\{x_1, \dots, x_i\}$, $i = 1, \dots, d$, the identity*

$$\det \big(f_i(x_i \wedge x_j)\big)_{i,j=1,\dots,d} = \prod_{i=1}^{d} \sum_{j=1}^{d} f_i(x_j)\mu(x_j, x_i) \tag{1.A.4}$$

holds, where μ is the Möbius function of the chain $\{x_1, \dots, x_d\}$, i.e.

$$\mu(x_j, x_i) = \begin{cases} 1 & i = j, \\ -1 & i = j+1, \\ 0 & \text{otherwise.} \end{cases} \tag{1.A.5}$$

Another proof that $R_{H,c}$ is positive definite. We show that the function

$$(s,t) \mapsto \begin{cases} \frac{(s \vee t)^\alpha}{(s \wedge t)^\beta} & s \wedge t > 0, \\ 0 & s \wedge t = 0, \end{cases} \tag{1.A.6}$$

is positive definite if and only if $\alpha + \beta \le 0$. Let $0 < t_1 \le t_2 \le \dots \le t_d$ and define

$$f_i(t_j) := (t_i/t_j)^{\alpha+\beta} = \frac{t_i^\beta (t_i \vee t_j)^\alpha}{(t_i \wedge t_j)^\beta t_j^\alpha}, \quad j \le i.$$

We then compute

$$\begin{aligned}
\det\left(\frac{(t_i \vee t_j)^\alpha}{(t_i \wedge t_j)^\beta}\right) &= (t_1 \cdots t_d)^{\alpha-\beta} \det\left(\frac{t_i^\beta (t_i \vee t_j)^\alpha}{(t_i \wedge t_j)^\beta t_j^\alpha}\right) \\
&= (t_1 \cdots t_d)^{\alpha-\beta} \prod_{i=1}^d \sum_{j=1}^d f_i(t_j)\mu(t_j, t_i) \\
&= t_d^{\alpha-\beta} \prod_{i=2}^d t_{i-1}^{\alpha-\beta}\Big(1 - \frac{t_i^{\alpha+\beta}}{t_{i-1}^{\alpha+\beta}}\Big) \\
&= t_d^{\alpha-\beta} \prod_{i=1}^{d-1} \frac{t_i^{\alpha+\beta} - t_{i+1}^{\alpha+\beta}}{t_i^{2\beta}},
\end{aligned} \tag{1.A.7}$$

where we have used (1.A.4) and (1.A.5). The statement about (1.A.6) now follows from the fact that a matrix is positive semidefinite if and only if all its principal minors have non-negative determinant.

As noted above, it is clear that the limiting covariance function $R_{H,-\infty}$ is also positive definite. This can be easily checked directly as well: Let $\{s_1, \dots, s_m\} = \{t_1, \dots, t_d\}$, where $t_1, \dots, t_d \geq 0$, and the s_i are distinct. Then (1.2.2) holds, since

$$\begin{aligned}
\sum_{k,l=1}^d a_k a_l R_{H,-\infty}(t_k, t_l) &= \sum_{i=1}^m \sum_{\substack{k,l=1 \\ t_k = t_l = s_i}}^d a_k a_l s_i^{2H} \\
&= \sum_{i=1}^m s_i^{2H} \Bigg(\sum_{\substack{k=1 \\ t_k = s_i}}^d a_k\Bigg)^2 \geq 0.
\end{aligned}$$

□

For $H > 0$ and $c < -H$, the process (1.2.1) has the Volterra representation

$$(X_t^{H,c})_{t\geq 0} \stackrel{\mathrm{d}}{=} \left(\int_0^t K_{H,c}(s,t)dB_s\right)_{t\geq 0},$$

where B is a Brownian motion, and

$$K_{H,c}(s,t) := \sqrt{-2(c+H)}\, t^{H-1/2}(s/t)^{-c-H-1/2}, \quad 0 \leq s \leq t.$$

Indeed, it is easily checked that this kernel satisfies

$$R_{H,c}(s,t) = \int_0^{s\wedge t} K_{H,c}(u,s)K_{H,c}(u,t)du, \quad s,t \geq 0.$$

Proposition 1.A.2. *Let $H > 0$ and $c = -H$. There is no function $F \in L^2(\mathbb{R}_+)$ such* *that*

$$(X^{H,-H})_{t\geq 0} \stackrel{\mathrm{d}}{=} \left(\int_0^t t^{H-1/2}F(u/t)dB_u\right)_{t\geq 0}.$$

Proof. This result follows from applying Theorem 2.1 and Remark 2.2 in [73] to the degenerate process $X_t^{H,-H} = t^H W(1)$, but can be easily proved directly as well: Suppose there is such a function F. By Ito's isometry, we have

$$(st)^H = R_{H,-H}(s,t) = \int_0^s t^{H-1/2}s^{H-1/2}F(u/t)F(u/s)du, \quad s \leq t,$$

and thus

$$\int_0^1 F(v)F(xv)dv = \frac{1}{\sqrt{x}}, \quad x > 0.$$

For $x = 1$, we obtain $\int_0^1 F(v)^2 dv = 1$. From this and the Cauchy-Schwarz inequality, we get

$$\begin{aligned}\frac{1}{x} &= \left(\int_0^1 F(v)F(xv)dv\right)^2 \\ &\leq \int_0^1 F(xv)^2 dv \int_0^1 F(v)^2 dv = \int_0^1 F(xv)^2 dv.\end{aligned}$$

Substituting $u = xv$ yields

$$\int_0^x F(u)^2 du \geq 1, \quad x \in (0,1],$$

and so $\int_0^1 F(u)^2 \mathbf{1}_{\{u>x\}} du = 0$. By monotone convergence,

$$0 = \lim_{x\downarrow 0} \int_0^1 F(u)^2 \mathbf{1}_{\{u>x\}} du = \int_0^1 F(u)^2 du,$$

which contradicts $\int_0^1 F(v)^2 dv = 1$. □

2. Self-similar Gaussian Markov processes

We characterize all multi-dimensional real self-similar Gaussian Markov processes. Three types of covariance matrix functions occur: white-noise type functions, covariances that can be expressed by continuous matrix semigroups, and covariances based on non-continuous solutions of Cauchy's functional equation. Characterizing the latter requires us to develop some results on the representation theory of non-continuous matrix semigroups, which are presented in chapter 3. The results in this chapter applied to the one dimensional case yields again the two-parameter family of time-changed Brownian motions discussed in chapter 1. This observation simplifies several proofs of non-Markovianity found in the literature.

2.1. Introduction

Our main contribution in this chapter is that we determine all *multi-dimensional* self-similar Gaussian Markov processes. By self-similarity, the bivariate covariance function can be reduced to a function of a single argument, and the functional equation mentioned above shows that this function must satisfy the semigroup property $g(x+y) = g(x)g(y)$. We identify a contractivity estimate involving the self-similarity parameter, which must be satisfied for self-similar Gaussian Markov processes with continuous semigroup. Under mild assumptions, *all* matrix solutions of the equation $g(x+y) = g(x)g(y)$, including non-measurable ones, are determined in chapter 3. While the self-similar Gaussian Markov processes resulting from these non-measurable semigroups are not likely to occur in applications, they have to be studied to obtain a full classification.

This chapter is organized as follows. In Section 2.2, we define the class of stochastic processes we wish to consider, and state our main characterization results. They are proven in Section 2.3, which includes a new criterion for checking positive definiteness in the multi-dimensional case. Section 2.4 presents Volterra representations for self-similar Gaussian Markov processes, assuming that the associated semigroup is continuous, and that the contractivity estimate mentioned above is strict. In Section 2.6 we show that the multidimensional characterization implies the statements in chapter 1. Section 2.7 returns to our initial motivation, by giving straightforward proofs that some self-similar Gaussian processes from the literature are not Markovian.

2.2. Preliminaries and main results

Throughout this section and the following one, V denotes a real d-dimensional vector space equipped with an inner product $\langle \cdot, \cdot \rangle$. We write $L(V)$ for the set of linear maps from V to itself. The inner product $\langle \cdot, \cdot \rangle_X$ associated with a V-valued Gaussian process X will be defined in Proposition 2.2.5, and we will write $\| \cdot \|_X$ for the corresponding operator norm on $L(V)$. The adjoint of a linear map will be denoted by *, where the inner product will be clear from the context, or explicitly stated otherwise.

Definition 2.2.1. *A stochastic process $(X_t)_{t\geq 0}$ with values in V is called Gaussian if, for any finite set $S \subset \mathbb{R}_{\geq 0}$, the $V^{|S|}$-valued random variable $(X_s)_{s\in S}$ is Gaussian.*

It is well known that a Gaussian process X with values in V is uniquely characterized in law by the two mappings $t \mapsto \mathbb{E}[X_t]$ and $(s,t) \mapsto R(s,t) \in L(V)$ given by

$$\langle v, R(s,t)u \rangle = \mathbb{E}[\langle X_s, v \rangle \langle X_t, u \rangle].$$

The mapping R is called the *covariance kernel* of X. We also write R_X, if X might not be clear from the context.

Definition 2.2.2. *For $H > 0$ we call X H-self-similar if $(X_{at})_{t\geq 0} \sim (a^H X_t)_{t\geq 0}$ (equality in distribution) for any $a > 0$, and refer to H as the self-similarity parameter.*

It is clear, and well-known, that H-self-similarity of a Gaussian process can be characterized via the covariance kernel.

Theorem 2.2.3. *Let X be a centered Gaussian process. Then X is H-self-similar if and only if R satisfies*

$$R(as, at) = a^{2H} R(s,t)$$

for any $a, s, t > 0$.

Corollary 2.2.4. *Let X be an H-self-similar centered Gaussian process on V. Then $\operatorname{supp}(X_t) = \ker(R(1,1))^{\perp}$ for all $t > 0$.*

Proof. By self-similarity, we have $\ker(R(1,1)) = \ker(R(t,t))$ for any $t > 0$. Denote by $p : V \to V$ the orthogonal projection onto $\ker(R(1,1))$ and set $Y_t := pX_t$. Clearly, Y is again Gaussian since the orthogonal projection is linear. Denote by R_Y the covariance kernel of Y. We have

$$\langle v_1, R_Y(s,t)v_2 \rangle = \mathbb{E}[\langle v_1, pX_s \rangle \langle v_2, pX_t \rangle], \quad v_1, v_2 \in V.$$

Since orthogonal projections are self-adjoint, we obtain

$$\mathbb{E}[\langle v_1, pX_s \rangle \langle v_2, pX_t \rangle] = \mathbb{E}[\langle pv_1, X_s \rangle \langle pv_2, X_t \rangle] = \langle v_1, pR(s,t)pv_2 \rangle,$$

where R is the covariance kernel of X. Since $pv_2 \in \ker(R(1,1))$, we obtain $R_Y(t,t) = 0$ for any $t \geq 0$. This easily implies that Y is the zero process. □

If $R(t,t) = t^{2H} R(1,1)$, $t > 0$, is not invertible, then the corollary shows that we are dealing with a self-similar Gaussian process of lower dimension, and V can be replaced by

$\ker(R(1,1))^{\perp}$. Thus, it suffices to characterize processes with full support. In this case, we can define the inner product

$$\langle v, u\rangle_X = \langle v, R(1,1)^{-1}u\rangle, \tag{2.2.1}$$

where $\langle \cdot, \cdot \rangle$ is the original inner product on V. This is well-defined, since $R(1,1)^{-1} \in L(V)$ is symmetric and positive definite.

Proposition 2.2.5. *Let X be a V-valued centered Gaussian process with $R(1,1)$ invertible. Then the inner product $\langle \cdot, \cdot\rangle_X = \langle \cdot, R(1,1)^{-1}\cdot\rangle$ is independent of the choice of $\langle \cdot, \cdot \rangle$.*

Proof. Let $\langle \cdot, \cdot \rangle$ be an inner product on V. For any other inner product on V there exists a symmetric positive map A such that it is given by $\langle \cdot, A\cdot\rangle =: \langle \cdot, \cdot \rangle_A$. Calculating the covariance kernel with respect to $\langle \cdot, \cdot \rangle_A$, we obtain

$$\begin{aligned}\mathbb{E}[\langle Av, X_s\rangle\langle Au, X_t\rangle] &= \langle Av, R(s,t)Au\rangle \\ &= \langle v, AR(s,t)Au\rangle = \langle v, R(s,t)Au\rangle_A, \quad u, v \in V.\end{aligned}$$

Hence the covariance kernel with respect to $\langle \cdot, \cdot \rangle_A$ is $R(s,t)A$. We calculate

$$\langle v, (R(1,1)A)^{-1}u\rangle_A = \langle v, A^{-1}R(1,1)^{-1}u\rangle_A = \langle v, R(1,1)^{-1}u\rangle.$$

Hence the definition of $\langle \cdot, \cdot \rangle_X$ does not depend on our choice of $\langle \cdot, \cdot \rangle$. ☐

In the notation of the preceding proof, the inner product (2.2.1) induced by X is an abbreviation for $\langle \cdot, \cdot \rangle_{R_X(1,1)^{-1}}$.

Lemma 2.2.6. *Let X be as in Proposition 2.2.5. Under $\langle \cdot, \cdot \rangle_X$, the covariance kernel of X is $R(s,t)R(1,1)^{-1}$.*

Proof. Put $A = R(1,1)^{-1}$ in the proof of the preceding proposition. ☐

Definition 2.2.7. *For $H > 0$, we call a V-valued process X an H-SSGM process, or just SSGM process, if*

(i) it is a centered H-self-similar Gaussian Markov process, and

(ii) it has full support, i.e., $R(1,1)$ is invertible.

For such a process, it is understood that the finite-dimensional vector space V is equipped with the inner product $\langle \cdot, \cdot \rangle_X$ and the corresponding operator norm. We write $\tilde{R}$, or sometimes $\tilde{R}_X$, for the covariance kernel w.r.t. $\langle \cdot, \cdot \rangle_X$, so that $\tilde{R}(1,1) = \mathrm{id}$ by Lemma 2.2.6.

The following well-known characterization of the Markov property for Gaussian processes is due to Doob [32, Theorem V.8.1]. The proof is usually presented for one-dimensional processes, but as the multivariate case is virtually the same, we omit it.

Lemma 2.2.8. *Let X be a centered V-valued Gaussian process such that $R(t,t)$ is invertible for all $t>0$. The process X satisfies the Markov property if and only if R satisfies the functional equation*

$$R(s,t)R(t,t)^{-1}R(t,u) = R(s,u), \quad 0 \le s \le t \le u. \tag{2.2.2}$$

If X is an SSGM process, then Lemma 2.2.6 implies that $\tilde{R}(t,t) = t^{2H}\,\mathrm{id}$ (recall the notation $\tilde{R}$ from Definition 2.2.7). Applying this to (2.2.2), we obtain

$$t^{-2H}\tilde{R}(s,t)\tilde{R}(t,u) = \tilde{R}(s,u),$$

and so

$$\tilde{R}(s/t,1)\tilde{R}(t/u,1) = \tilde{R}(s/u,1).$$

Substituting $x := \log(t/s)$ and $y := \log(u/t)$ and setting $g(x) := \tilde{R}(e^{-x},1)$ yields

$$g(x)g(y) = g(x+y), \quad x, y \ge 0, \tag{2.2.3}$$

and hence g is a one-parameter semigroup.

Definition 2.2.9. *A map $g : \mathbb{R}_{\ge 0} \to L(V)$ is a semigroup if $g(0) = \mathrm{id}$ and $g(x+y) = g(x)g(y)$ for all $x, y \ge 0$.*

Definition 2.2.10. *For any SSGM process X, we call the one-parameter semigroup $(g_X(x))_{x\ge 0}$, where $g_X(x) := \tilde{R}(\exp(-x),1)$, its associated semigroup. If there is no ambiguity, we just write g instead of g_X.*

We stress that our notion of associated semigroup, defined by the scaled covariance function, is *not* the transition semigroup of the Markov process, which will not be used here. The following theorem, proven in Section 2.3, is our first main result.

Theorem 2.2.11. *Let $H > 0$, let $(g(x))_{x\ge 0}$ be a semigroup in $L(V)$, and $\langle \cdot, \cdot \rangle$ an inner product on V. There exists an SSGM process X with associated semigroup $g = g_X$ and inner product $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_X$ if and only if the contractivity estimate*

$$\|g(x)\|_{\mathrm{op}} \le e^{-Hx}, \quad x \ge 0, \tag{2.2.4}$$

holds, where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm of $\langle \cdot, \cdot \rangle$. Furthermore, X is determined uniquely in law by g_X and $\langle \cdot, \cdot \rangle_X$.

Definition 2.2.12. *We call an SSGM process white noise if its associated semigroup satisfies $g(x) = 0$ for all $x > 0$.*

By definition of g, the covariance kernel of a white noise process satisfies $\tilde{R}(s,t) = 0$ for $0 < s < t$.

Definition 2.2.13. *Let X be an SSGM process on V. We say that X is irreducible if for any decomposition $X = X^1 + X^2$ such that X^1 and X^2 are independent SSGM processes supported on V_1 and V_2 respectively with $V_1 \perp V_2$, the spaces V_1 and V_2 are trivial, i.e. either the entire space or $\{0\}$. The orthogonality of the subspaces is with respect to $\langle \cdot, \cdot \rangle_X$.*

Remark 2.2.14. *As $\langle \cdot, \cdot \rangle_X$ depends on X only through its law, so does the condition that the joint law of (X^1, X^2) must satisfy in the preceding definition. Therefore, assuming only equality in distribution, $X \sim X^1 + X^2$, yields an equivalent definition.*

We proceed to characterize all semigroups associated with centered self-similar Gaussian Markov processes. Consider Cauchy's functional equation

$$f(x) + f(y) = f(x+y), \quad f : \mathbb{R} \to \mathbb{R}. \tag{2.2.5}$$

Define the rotation matrix

$$Q(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}, \quad \theta \in \mathbb{R},$$

and, for even k, the block-diagonal matrix

$$Q_k^\nu(x) := \underbrace{\begin{pmatrix} Q(\nu(x)) & & 0 \\ & \ddots & \\ 0 & & Q(\nu(x)) \end{pmatrix}}_{k \times k} \tag{2.2.6}$$

consisting of $k/2$ rotation matrices evaluated at some solution ν of (2.2.5). Now we state our second main theorem, which gives representations for the covariance kernels of all irreducible H-self-similar Markov processes. Recall the notation defined in (2.2.1), Definition 2.2.10, and (2.2.6).

Theorem 2.2.15. *Let X be an irreducible SSGM process with values in V. Then there exists an orthonormal basis with respect to $\langle \cdot, \cdot \rangle_X$ such that with respect to this basis one of the following holds (we identify linear maps and matrices if the chosen basis is clear from the context; see Remark 2.2.16 for details):*

(i) $g_X(x) = 0$ *for* $x > 0$ *and* V *is one-dimensional (white noise case),*

(ii) $g_X(x) = \exp(Mx)$, $x \geq 0$, *for some matrix* M *(the generator), or*

(iii) The dimension d *is even, and* $g_X(x) = Q_d^\nu(x) \exp(Mx)$, $x \geq 0$, *with* ν *a non-continuous solution to Cauchy's functional equation (2.2.5) and* M *commuting with every* $Q_d^\nu(x)$.

In the second and third case we have the contractivity condition $\|\exp(Mx)\|_X \leq e^{-Hx}$, $x \geq 0$, where $\|\cdot\|_X$ is the operator norm on $L(V)$ induced by $\langle \cdot, \cdot \rangle_X$.

The theorem is proven at the end of Section 2.3.

Remark 2.2.16. *Let $\Gamma : V \to \mathbb{R}^d$ be an isomorphism that maps vectors to their coordinates w.r.t. to the orthonormal basis from the preceding theorem. If $\tilde{A}$ denotes the matrix of the inner product $\langle \cdot, \cdot \rangle_X$ in these coordinates, i.e. $\langle \Gamma^{-1} w, \Gamma^{-1} z \rangle_X = w^{\mathrm{T}} \tilde{A} z$, $w, z \in \mathbb{R}^d$,* *then the identification of linear maps and representation matrices we used in the theorem means that* $\|\exp(Mx)\|_X = \|\exp(Mx)\|_{\tilde{A}}$*, where* $\|\cdot\|_{\tilde{A}}$ *is the operator norm induced by the inner product* $(w, z) \mapsto w^{\mathrm{T}}\tilde{A}z$ *on* $\mathbb{R}^d$*. Similarly, in the following remark the expression on the left hand side of* (2.2.7) *is an abbreviation for* $(\Gamma v)^{\mathrm{T}}(M + H\,\mathrm{id})^{\mathrm{T}}\tilde{A}\Gamma v$*.*

Remark 2.2.17. *The contractivity condition* $\|\exp(Mx)\|_X \le e^{-Hx}$*,* $x \ge 0$*, is equivalent to the dissipativity of the (in general non-symmetric) matrix* $M + H\,\mathrm{id}$*, i.e.,*

$$\langle (M + H\,\mathrm{id})v, v\rangle_X \le 0, \quad v \in V. \tag{2.2.7}$$

This follows from the Lumer–Phillips theorem [27, p. 52], which characterizes contractive semigroups. In the special case where $\tilde{A} = \mathrm{id}$ *and the matrix* M *consists of a single Jordan block, this condition can be simplified further. See chapter 4 for details, where we also point out a relation to a classical inequality due to Fan–Taussky–Todd [36].*

Using Theorem 2.2.15, we can give the associated semigroup of any SSGM process X. By the definition of irreducibility, there exists an orthogonal decomposition $V = \bigoplus_{i=1}^n V_i$ and independent irreducible SSGM processes X^i supported on V_i respectively such that $X = \sum_{i=1}^n X^i$ and $g_X = \bigoplus_{i=1}^n g_{X^i}$. By Theorem 2.2.15, each g_{X^i} has one of the forms given in Theorem 2.2.15 and satisfies the contractivity condition $\|g_{X^i}(x)\|_{X^i} \le e^{-Hx}$. Conversely, given semigroups $g_1, \dots, g_n$ on finite-dimensional vector spaces $V_1, \dots, V_n$ such that each g_i is of the form given in Theorem 2.2.15 and satisfies the contractivity condition, then by Theorem 2.2.11, for any inner product $\langle\cdot,\cdot\rangle$ on $\bigoplus_{i=1}^n V_i$ such that the V_i are orthogonal, there exists a unique in law SSGM process X with associated semigroup $\bigoplus_{i=1}^n g_i$.

2.3. Proofs

As in Section 2.2, V denotes a real d-dimensional vector space with inner product $\langle\cdot,\cdot\rangle$. For one-dimensional Gaussian processes, the following result is standard. The proof in dimension d is a straightforward extension; see [21, Theorem 2].

Theorem 2.3.1. *Let* $R : \mathbb{R}_{\ge 0} \times \mathbb{R}_{\ge 0} \to L(V)$*. Then there exists a centered Gaussian process with covariance kernel* R *iff*

(i) $R(s,t) = R(t,s)^*$ *for* $s, t \ge 0$*, where the adjoint is taken with respect to* $\langle\cdot,\cdot\rangle$*,*

(ii) for all $n > 0$ *we have* $\sum_{i,j=1}^n \langle v_i, R(t_i,t_j)v_j\rangle \ge 0$ *for* $v_1, \dots, v_n \in V$ *and* $t_1, \dots, t_n \in \mathbb{R}_{\ge 0}$*.*

This process is unique in law.

The following characterization of (ii) in Theorem 2.3.1 seems to be new.

Lemma 2.3.2. *Let $R : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to L(V)$ be a kernel which satisfies (2.2.2), $R(t,t)$ being symmetric and positive definite for all $t \geq 0$ and $R(s,t)^* = R(t,s)$ for all $s,t \geq 0$. Then R satisfying condition (ii) in Theorem 2.3.1 is equivalent to*

$$R(s,s) - R(s,t)R(t,t)^{-1}R(t,s)$$

being positive semidefinite for all $s,t \in \mathbb{R}_{\geq 0}$.

Proof. We first show necessity. Clearly it suffices to consider the problem in $\mathbb{R}^n$ equipped with the standard inner product, by applying the isomorphism Γ from Remark 2.2.16. Notice that condition (ii) in Theorem 2.3.1 is equivalent to

$$R^n = \begin{pmatrix} R(t_1,t_1) & \dots & R(t_1,t_n) \\ \vdots & \ddots & \\ R(t_n,t_1) & & R(t_n,t_n) \end{pmatrix} \tag{2.3.1}$$

being positive semi-definite for all $t_1, \dots, t_n \geq 0$ and $n > 0$. For every $s,t \geq 0$ denote by $A_{s,t}$ a solution of

$$A_{s,t}^{\mathrm{T}} A_{s,t} = R(t,t) - R(t,s)R(s,s)^{-1}R(s,t).$$

Let $M^n(L(V))$ the algebra of $n \times n$ matrices over the space $L(V)$. Clearly we have $R^n \in M^n$. Notice that each $M \in M^n(L(V))$ defines a linear map on V^n. Denote by $M^{\mathrm{T}} \in M^n(L(V))$ the transpose of the operator M on V^n. We show by induction that the covariance operator (2.3.1) is a positive semidefinite operator on V^n for any $n \in \mathbb{N}$ and any increasing sequence $t_1, \dots, t_n \in \mathbb{R}_{\geq 0}$. By assumption, R^1 is positive semidefinite. Assume that R^n is positive semidefinite. We define

$$R^{n,n+1} := \begin{pmatrix} R^n & 0_n^{\mathrm{T}} \\ 0_n & \mathrm{id} \end{pmatrix},$$

where $0_n = (0, \dots, 0)$ with $0 : V \to V$ mapping everything to zero. It follows that $R^{n,n+1}$ is positive semidefinite on $\mathbb{R}^{d(n+1)}$. Let

$$B_{n+1} := \begin{pmatrix} \mathrm{id}_n & v_{n+1}^{\mathrm{T}} \\ 0 & A_{t_n,t_{n+1}} \end{pmatrix},$$

where id_n is the identity in $M^n(L(V))$ and

$$v_{n+1} := \big(0, \dots, 0, R(t_n,t_n)^{-1}R(t_n,t_{n+1})\big) \in L(V)^{1 \times n}.$$

Then we have

$$B_{n+1}^{\mathrm{T}} R^{n,n+1} B_{n+1} = R^{n+1}.$$

This shows that R^{n+1} is positive semidefinite, finishing the induction step. Conversely, by using condition (ii) in Theorem 2.3.1 for $n = 2$, we have

$$\langle v_1, R(s,s)v_1 \rangle - 2\langle v_1, R(s,t)v_2 \rangle + \langle v_2, R(t,t)v_2 \rangle \geq 0.$$

Setting $v_2 = R(t,t)^{-1}R(t,s)v_1$ yields

$$\langle v_1, R(s,s)v_1 - R(s,t)R(t,t)^{-1}R(t,s)v_1\rangle \geq 0,$$

which shows that $R(s,s) - R(s,t)R(t,t)^{-1}R(t,s)$ is positive semidefinite. $\square$

Combining Theorem 2.3.1, Lemma 2.2.8 and Lemma 2.3.2 yields the following result. The new aspect of Theorem 2.3.3 is that (iii), together with the other conditions, implies that R is a covariance kernel.

Theorem 2.3.3. *Let $R : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to L(V)$ be such that $R(t,t)$ is invertible for all $t > 0$. Then there exists a centered Gaussian Markov process with covariance kernel R if and only if*

(i) $R(t,t)$ is positive definite for all $t > 0$,

(ii) $R(s,t) = R(t,s)^*$ for any $s,t \geq 0$, where the adjoint is taken with respect to $\langle \cdot,\cdot \rangle$,

(iii) $R(s,s) - R(s,t)R(t,t)^{-1}R(t,s)$ is positive semidefinite for all $s,t \geq 0$,

(iv) $R(s,t)R(t,t)^{-1}R(t,u) = R(s,u)$ for all $0 \leq s < t < u$ (Markov property).

The following theorem motivates Definition 2.2.13, as it tells us that the associated semigroup of any reducible SSGM process is itself reducible in an appropriate sense.

Theorem 2.3.4. *An SSGM process is irreducible if and only if there does not exist a nontrivial subspace $W \subset V$ such that $g_X(x)W = W$ and $g_X(x)W^\perp = W^\perp$ for all $x \geq 0$.*

Proof. Assume first that X is *not* irreducible, i.e. $X = X^1 + X^2$ and $V = W \oplus W^\perp$, with X^1 supported on W and X^2 supported on $W^\perp$. By assumption X^1 is independent of X^2. Denote by p_1 the orthogonal projection onto W and p_2 the orthogonal projection onto $W^\perp$. Since $R_{X^i}(1,1) = p_i R_X(1,1) p_i = p_i$ which is the identity on W and $W^\perp$ respectively, we have that $\langle\cdot,\cdot\rangle_{X^1}$ and $\langle\cdot,\cdot\rangle_{X^2}$ are just the restrictions of $\langle\cdot,\cdot\rangle_X$ to $W \times W$ and $W^\perp \times W^\perp$ respectively. Denote by $g_{X^1}(x) : W \to W$ and $g_{X^2}(x) : W^\perp \to W^\perp$ the associated semigroups of X^1 and X^2 respectively. We show that $g_X(x) = g_{X^1}(x)p_1 + g_{X^2}(x)p_2$. Clearly, for any $v \in V$ we have $v = p_1 v + p_2 v$ and hence also $p_i X = X^i$. Furthermore $p_1^* = p_1$ and $p_2^* = p_2$. For any $v, u \in V$ we obtain

$$\begin{aligned}\langle v, g_X(x)u\rangle_X &= \langle v, \tilde{R}_X(e^{-x},1)u\rangle_X = \mathbb{E}[\langle v, X_{e^{-x}}\rangle_X \langle u, X_1\rangle_X] \\ &= \mathbb{E}\big[\langle v, X^1_{e^{-x}} + X^2_{e^{-x}}\rangle_X \langle u, X^1_1 + X^2_1\rangle_X\big].\end{aligned}$$

Since X^1 is independent of X^2 we have

$$\mathbb{E}[\langle v, X^1_{e^{-x}}\rangle_X \langle u, X^2_1\rangle_X] = \mathbb{E}[\langle v, X^2_{e^{-x}}\rangle_X \langle u, X^1_1\rangle_X] = 0,$$

and hence

$$\begin{aligned}\mathbb{E}\big[\langle v, X^1_{e^{-x}} + X^2_{e^{-x}}\rangle_X \langle u, X^1_1 + X^1_2\rangle_X\big] &= \mathbb{E}\big[\langle v, X^1_{e^{-x}}\rangle_X \langle u, X^1_1\rangle_X + \langle v, X^2_{e^{-x}}\rangle_X \langle u, X^2_1\rangle_X\big] \\ &= \mathbb{E}\big[\langle p_1 v, X^1_{e^{-x}}\rangle_X \langle p_1 u, X^1_1\rangle_X + \langle p_2 v, X^2_{e^{-x}}\rangle_X \langle p_2 u, X^2_1\rangle_X\big] \\ &= \langle p_1 v, g_{X^1}(x) p_1 u\rangle_X + \langle p_2 v, g_{X^2}(x) p_2 u\rangle_X \\ &= \langle v, (g_{X^1}(x) p_1 + g_{X^2}(x) p_2) u\rangle_X.\end{aligned}$$

Conversely, assume that there exists a nontrivial subspace W such that W and $W^\perp$ are invariant under $g_X(x)$. We show that $X^1 := p_1 X$ and $X^2 := p_2 X$ are independent SSGM processes. Clearly, $X = X^1 + X^2$, and since p_1, p_2 are both linear, X^1 and X^2 are H-self-similar Gaussian. It remains to show that they are independent and Markovian. For Gaussian processes it is enough to show pointwise independence. For $0 \le s \le t$ and $v, u \in V$ we have

$$\begin{aligned}\mathbb{E}[\langle v, X^1_s\rangle_X \langle u, X^2_t\rangle_X] &= \mathbb{E}[\langle p_1 v, X_s\rangle_X \langle p_2 u, X_t\rangle_X] \\ &= \langle p_1 v, \tilde{R}_X(s,t) p_2 u\rangle_X = t^2 H \langle p_1 v, g_X(\log(t/s)) p_2 u\rangle_X = 0,\end{aligned}$$

where the last equality follows from $p_1 v \in W$ and $g_X(\log(t/s)) p_2 u \in W^\perp$. In the same manner we obtain independence for $s > t$. For any $v, u \in V$ we have

$$\mathbb{E}[\langle v, X^1_{e^{-x}}\rangle_X \langle u, X^1_1\rangle_X] = \mathbb{E}[\langle p_1 v, X_{e^{-x}}\rangle_X \langle p_1 u, X_1\rangle_X] = \langle v, p_1 g_X(x) p_1 u\rangle_X.$$

Hence $g_{X^1}(x) = p_1 g_X(x) p_1$. Analogously, we get $g_{X^2}(x) = p_2 g_X(x) p_2$. It follows that $g_{X^1}(x)$ and $g_{X^2}(x)$ are semigroups, and hence X^1 and X^2 satisfy condition (iv) in Theorem 2.3.3. $\square$

We can now prove our main results.

Proof of Theorem 2.2.11. Sufficiency: Define

$$\tilde{R}(s,t) = t^{2H} g(\log(t/s)), \quad 0 < s \le t,$$

and set $\tilde{R}(s,t) = \tilde{R}(t,s)^*$ for $s > t$, where the adjoint is with respect to $\langle\cdot,\cdot\rangle$. We now apply Theorem 2.3.3 to $\tilde{R}$, where V is equipped with the inner product $\langle\cdot,\cdot\rangle$. Assumption (i) follows from Definition 2.2.9, as does (iv), using also the calculation above (2.2.3). Assumption (ii) is clear by definition of $\tilde{R}$, and (iii) easily reduces to

$$\|g(x)\|_{\mathrm{op}} \le e^{-Hx}, \quad x \ge 0. \tag{2.3.2}$$

Hence there exists a Gaussian Markov process X with covariance kernel $\tilde{R}$. The resulting process X is self-similar by Theorem 2.2.3. The uniqueness assertion follows from Theorem 2.3.1.

Necessity: Let R be the covariance kernel of X w.r.t. $\langle\cdot,\cdot\rangle$. As in Definition 2.2.10, we put $g_X(x) = \tilde{R}(e^{-x}, 1)$. It is easy to see that part (iii) of Theorem 2.3.3 shows that (2.2.4) holds for $\langle\cdot,\cdot\rangle_X = \langle\cdot, R(1,1)^{-1}\cdot\rangle$. $\square$

The following characterization of locally bounded matrix semigroups is proven in chapter 3. It holds for V as at the beginning of Section 2.2, with an arbitrary inner product $\langle\cdot,\cdot\rangle$, and we write $\|\cdot\|$ for the corresponding operator norm. Below, the theorem will be applied to $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_X$.

Theorem 2.3.5. *Let $(g(x))_{x\geq 0}$ be a semigroup in $L(V)$ such that $\|g(x)\| \leq f(x)$ for all $x \geq 0$ with f being locally bounded and right-continuous at 0 with $f(0) = 1$. Then there exists an orthogonal decomposition $V = \bigoplus_{i=1}^n V_i$ such that each V_i is invariant under $g(x)$ and either $g(x)$ is elementary on V_i, or $g(x)|_{V_i} = 0$ for $x > 0$. Here,* elementary *means that the semigroup is of the form (ii) or (iii) in Theorem 2.2.15.*

Proof of Theorem 2.2.15. Let X be an irreducible SSGM process. By Theorem 2.3.4, V does not have a decomposition into nontrivial orthogonal subspaces which are invariant under g_X. By Theorem 2.2.11, we have $\|g_X(x)\|_X \leq e^{-Hx}$, and hence we can apply Theorem 2.3.5. Because of the irreducibility this decomposition has to be trivial, and hence either $g_X(x)$ is elementary on V or $g_X(x) \equiv 0$. If $g_X(x) \equiv 0$ for $x > 0$ then X is white noise. In this case, for any decomposition $V = W \oplus W^\perp$ the projections $p_W(X)$ and $p_{W^\perp}(X)$ are again H-self-similar white noise. Since they are white noise they are also Markovian. In order for X to be irreducible any such decomposition has to be trivial and hence V must be one-dimensional. $\square$

2.4. Volterra representation

We now assume $V = \mathbb{R}^d$, equipped with the standard inner product, and our goal is to give a Volterra representation of a continuous centered SSGM process X with semigroup $(\exp(Mx))_{x\geq 0}$ satisfying

$$\|e^{Hx}\exp(Mx)\|_X < 1, \quad x > 0. \tag{2.4.1}$$

Thus, we consider the non-degenerate case (ii) of Theorem 2.2.15, and the bound at the end of that theorem is assumed to be strict. We will show that there is a (deterministic) kernel $K(s,t)$ such that X equals $\int_0^\cdot K(s,\cdot)dW_s$ in distribution, where W is a Brownian motion; this is commonly referred to as a Volterra representation of the process. Under the standard inner product, the covariance kernel of X takes the form

$$R(s,t) = t^{2H}\exp(M\log(t/s))A^{-1}, \quad 0 < s < t, \tag{2.4.2}$$

where $A := R(1,1)^{-1}$ is a positive definite real matrix satisfying

$$\langle u, v\rangle_X = u^{\mathrm{T}}Av, \quad u, v \in \mathbb{R}^d.$$

Lemma 2.4.1. *Under the above assumptions, the matrix $M^{\mathrm{T}}A + AM + 2HA$ is negative definite.*

Proof. By the Lumer–Phillips theorem (see Remark 2.2.17), the bound $\|\exp((M + H\,\mathrm{id})x)\|_X \leq 1$ shows that $\langle (M + H\,\mathrm{id})v, v\rangle_X \leq 0$. By Theorem 4.4.1 in chapter 4, the strict bound (2.4.1) implies the stronger assertion

$$\langle (M + H\,\mathrm{id})v, v\rangle_X < 0, \quad 0 \neq v \in \mathbb{R}^d.$$

(While the standard inner product is considered in chapter 4, the proof of the theorem we have just used easily extends to an arbitrary inner product on $\mathbb{R}^d$.) We have, by symmetry of the inner product,

$$\begin{aligned} 2\langle (M + H\,\mathrm{id})v, v\rangle_X &= \langle (M + H\,\mathrm{id})v, v\rangle_X + \langle v, (M + H\,\mathrm{id})v\rangle_X \\ &= v^{\mathrm{T}}(M^{\mathrm{T}}A + AM + 2HA)v. \end{aligned}$$

□

Lemma 2.4.2. *Given a $d \times d$ matrix Q with $\lim_{x\to\infty} \exp(Qx) = 0$ and a $d \times d$ matrix P, we have, in the sense of the Cauchy principal value,*

$$P = -\int_0^\infty \exp(Q^{\mathrm{T}}x)(Q^{\mathrm{T}}P + PQ)\exp(Qx)dx.$$

Proof. We have $\frac{d}{dx}\exp(Q^{\mathrm{T}}x)P\exp(Qx) = \exp(Q^{\mathrm{T}}x)(Q^{\mathrm{T}}P + PQ)\exp(Qx)$, and hence

$$\int_0^t \exp(Q^{\mathrm{T}}x)(Q^{\mathrm{T}}P + PQ)\exp(Qx)dx = \exp(Q^{\mathrm{T}}t)P\exp(Qt) - P.$$

We conclude by taking the limit $t \to \infty$. □

The following theorem is the main result of this section.

Theorem 2.4.3. *Let X, H, M, and A be as at the beginning of this section. In particular, (2.4.1) holds. Define*

$$K(s,t) = t^{H-1/2}A^{-1}\exp(S\log(t/s))U, \quad 0 < s < t,$$

where U is a positive definite solution to $U^2 = -(M^{\mathrm{T}}A + AM + 2HA)$, and $S = M^{\mathrm{T}} + (H + 1/2)\,\mathrm{id}$. Then the centered Gaussian process defined by

$$\tilde{X}_t = \int_0^t K(s,t)dW_s, \quad t \geq 0,$$

where $(W_t)_{t\geq 0}$ is a standard d-dimensional Brownian motion, has covariance kernel (2.4.2). Therefore, it equals X in distribution.

Proof. Note that U is well-defined by Lemma 2.4.1. We assume $0 < s \leq t$ throughout the proof. Clearly, we have

$$R_{\tilde{X}}(s,t) = \mathbb{E}[\tilde{X}_s(\tilde{X}_t)^{\mathrm{T}}] = \int_0^s K(u,s)K(u,t)^{\mathrm{T}}du,$$

where the second equality follows from Ito's isometry. From $K(s,t) = t^{H-1/2}K(s/t,1)$

we obtain

$$\int_0^s K(u,s)K(u,t)^{\mathrm{T}}du = (st)^{H-1/2}\int_0^s K(u/s,1)K(u/t,1)^{\mathrm{T}}du,$$

and substituting $u = sr$ yields

$$(st)^{H-1/2}\int_0^s K(u/s,1)K(u/t,1)^{\mathrm{T}}du = s^{H+1/2}t^{H-1/2}\int_0^1 K(r,1)K(rs/t,1)^{\mathrm{T}}dr.$$

Now

$$K(r,1)K(rs/t,1)^{\mathrm{T}} = A^{-1}\exp(S\log(1/r))U^2\exp(S^{\mathrm{T}}\log(1/r))\exp(S^{\mathrm{T}}\log(t/s))A^{-1}.$$

Hence

$$\begin{aligned}
&s^{H+1/2}t^{H-1/2}\int_0^1 K(r,1)K(rs/t,1)^{\mathrm{T}}dr\\
&\quad= s^{H+1/2}t^{H-1/2}A^{-1}\int_0^1 \exp(S\log(1/r))U^2\exp(S^{\mathrm{T}}\log(1/r))dr\exp(S^{\mathrm{T}}\log(t/s))A^{-1}\\
&\quad= A^{-1}\left[\int_0^\infty \exp((M^{\mathrm{T}}+H\,\mathrm{id})x)U^2\exp((M+H\,\mathrm{id})x)dx\right]\ t^{2H}\exp(M\log(t/s))A^{-1},
\end{aligned}$$

where we made the substitution $r = e^{-x}$. It suffices to calculate the integral

$$\int_0^\infty \exp((M^{\mathrm{T}}+H\,\mathrm{id})x)U^2\exp((M+H\,\mathrm{id})x)dx.$$

Since $\|\exp((H\,\mathrm{id}+M)x)\|_X < 1$ for $x > 0$ by (2.4.1), we have the bound

$$\|\exp((H\,\mathrm{id}+M)x)\|_X < (\|\exp((H\,\mathrm{id}+M))\|_X)^n, \tag{2.4.3}$$

where $n = \lfloor x \rfloor$. Clearly this converges to 0 as $n \to \infty$. Setting $Q := M + H\,\mathrm{id}$ and $P := A$ in Lemma 2.4.2, we obtain

$$\int_0^\infty \exp((M^{\mathrm{T}}+H\,\mathrm{id})x)U^2\exp((M+H\,\mathrm{id})x)dx = A. \tag{2.4.4}$$

Note that the integral exists, not just as a Cauchy principal value, by the estimate (2.4.3). Rewriting (2.4.4) yields

$$\int_0^s K(u,s)K(u,t)^{\mathrm{T}}du = t^{2H}\exp(M\log(t/s))A^{-1}. \qquad \square$$

2.5. Impossibility of Volterra representations

In this Section we expand on section 2.4 to discuss under what condition it is impossible to give a Volterra representation. Let X be a continuous irreducible centered SSGM process with self similarity coefficient H. The covariance kernel is of the form

$$R(s,t) = t^{2H} \exp(M \log(t/s))$$

for $t \geq s$ under the intrinsic inner product $\langle \cdot, \cdot \rangle_X$. The space $W_M = \{v | \; \forall x \geq 0 : ||\exp(Mx)v|| = e^{-Hx} \; \}$ and its orthogonal complement are invariant under the semi group action of $\exp(Mx)$ (see Lemma 4.4.4) and hence invariant under the covariance kernel. Since X is irreducible either W_M or $W_M^{\perp}$ is the entire space.

The ability to give a Volterra representation of a self similar process is equivalent to the right limit of its filtration at time 0 being trivial, i.e. $\cap_{t\geq 0}\mathcal{F}_t^X$ as shown in [43]. Equivalently by defining the set of random variables

$$H_X(t) := \overline{\operatorname{span}\{\langle X_s, v\rangle | s \leq t, \;\; v \in V\}} \tag{2.5.1}$$

where the closure is taken in $L^2(\mathbb{P})$ with $\mathbb{P}$ being the measure induced by X, we have the following theorem, which is Theorem 2.1 in [74]:

Theorem 2.5.1. *The centered self similar Gaussian process X admits a Volterra representation if and only if $\bigcap_{t>0} H_X(t)$ is the $L^2(\mathbb{P})$ subspace spanned by constant functions.*

Processes satisfying (2.5.1) are called purely non-deterministic. In contrast any process for which the sets $H_X(t)$ remain constant for $t > 0$ or equivalently the filtration remains constant for $t > 0$ is called deterministic.

Lemma 2.5.2. *Let X be an irreducible continuous centered SSGM process with self similarity coefficient H. If $||\frac{1}{(ts)^H}R(s,t)||_{\mathrm{op}} = 1$ for some fixed $0 < s < t$ and thus for any $t \geq s > 0$, then X is deterministic.*

Proof. We have $R(s,t) = t^{2H}\exp(M\log(t/s))$ for $0 < s \leq t$. In order to show that X is deterministic we show that $H_X(t) = H_X(s)$ for all $0 < s < t$. Since $||\frac{1}{(ts)^H}R(s,t)||_{\mathrm{op}} = 1$ we have that W_M is the entire space and thus $\exp([M + H\,\mathrm{id}]x)$ is an isometry, i.e. for any w there exists v such that $\exp(M\log(t/s))w = e^{-H\log(t/s)}v$ and $||w|| = ||v||$. This is equivalent to $R(s,t)w = (st)^H v$. We have

$$\underbrace{\left\langle \left(\frac{t}{s}\right)^H v, \left(\frac{t}{s}\right)^H R(s,s)v \right\rangle}_{=t^{2H}} + \underbrace{\left\langle w, R(t,t)w \right\rangle}_{=t^{2H}} - 2\underbrace{\left\langle \left(\frac{t}{s}\right)^H v, R(s,t)w \right\rangle}_{=t^{2H}} = 0.$$

The above expression is equivalent to

$$\mathbb{E}\Big[\Big(\Big\langle X_t, w\Big\rangle - \Big\langle X_s, \left(\frac{t}{s}\right)^H v\Big\rangle\Big)^2\Big] = 0$$

and hence we have $\left\langle X_t, w \right\rangle = \left\langle X_s, \left(\frac{t}{s}\right)^H v \right\rangle$ almost surely. Since w was arbitrary we obtain $H_X(t) \subset H_X(s)$ and thus $H_X(t) = H_X(s)$. It follows that X is deterministic and hence does not admit a Volterra representation. $\square$

Remark 2.5.3. *If one assumes $X_t = \int_0^t k(s,t)dW_s$ for some L^2 integrable kernel then we obtain $H_X(t) \subset H_W(t)$ and hence X must satisfy the condition of Theorem 2.5.1. Hence one can see that it is impossible for a deterministic process to have a Volterra representation.*

Corollary 2.5.4. *Let X be an irreducible continuous centered SSGM process with self similarity coefficient $H > 0$. If $||\frac{1}{(ts)^H} R(s,t)||_{\text{op}} < 1$ for any $t \geq s > 0$ then X is purely non-deterministic.*

Proof. Since $||\frac{1}{(ts)^H} R(s,t)||_{\text{op}} < 1$ and X is irreducible we obtain

$$|| \exp(Mx)v|| < e^{-Hx}$$

for all $v \neq 0$ and $x > 0$. By Theorem 2.4.3 X admits a Volterra representation and hence it is purely non-deterministic. $\square$

Theorem 2.5.5. *If X is an irreducible continuous centered SSGM process then it is deterministic if $e^{Hx}g_X(x)$ is a semigroup of isometries and hence there does not exists a Volterra representation. Otherwise there exists a Volterra representation and X is purely non-deterministic.*

Proof. This follows from Corollary 2.5.4 and Lemma 2.5.2 and the fact that due to the irreduciblity of X either W_M or its orthogonal complement is the entire space. Theorem 2.5.1 implies the non existence of a Volterra representation in the deterministic case. If X is non-deterministic, then by Lemma 2.5.1 the semigroup $e^{Hx}g_X(x)$ satisfies $||e^{Hx}g_X(x)||_{\text{op}} < 1$ and thus by Theorem 2.4.3 there exists a Volterra representation and thus X is purely non-deterministic. $\square$

2.6. One-dimensional processes

Specializing our multi-dimensional characterization results, Theorem 2.2.11 and Theorem 2.2.15, we easily easily recover Theorem 1.2.3:

Theorem 2.6.1. *If X is an $\mathbb{R}$-valued H-SSGM process which is not white noise (see Definitions 2.2.7 and 2.2.12), then*

$$(X_t)_{t>0} \sim (at^{2H+M}W(t^{-2H-2M}))_{t>0},$$

where $M \in \mathbb{R}$, $a = R(1,1)^{1/2} > 0$, $H + M \leq 0$ and $(W(t))_{t\geq 0}$ is a standard Brownian motion. The covariance function of X is $R(s,t) = a^2 t^{2H}(t/s)^M$, $0 < s \leq t$.

Proof. Since X is one-dimensional, it has to be irreducible. As X is not white noise, Theorem 2.2.15 yields $g_X(x) = e^{Mx}$ for some $M \in \mathbb{R}$. We obtain $\tilde{R}(s,t) = t^{2H}(t/s)^M$ for $t \geq s > 0$; recall that $\tilde{R}$ is the covariance kernel with respect to the inner product $\langle \cdot, \cdot \rangle_X$. By multiplying with $R(1,1) =: a^2$, we obtain $R(s,t) = a^2 t^{2H}(t/s)^M$. It remains to check the contractivity condition on g_X. By Theorem 2.2.11 we must have $\|g_X(x)\|_X \leq e^{-Hx}$. Since $\langle x, y \rangle_X = a^{-2}xy$, we obtain $\|g_X(x)\|_X = e^{Mx}$. It follows that $H + M \leq 0$ (cf. Remark 2.2.17). One can easily check that $at^{2H+M}W(t^{-2H-2M})$ has the same covariance kernel. $\square$

The pointwise limit

$$\lim_{M \to -\infty} a^2 t^{2H}(t/s)^M = \begin{cases} a^2 t^{2H} & s = t, \\ 0 & s < t, \end{cases}$$

of the covariance function $R(s,t) = a^2 t^{2H}(t/s)^M$ is positive definite as well, and defines a white noise process (for $a > 0$), which is also self-similar and Markovian. This corresponds to case (i) in Theorem 2.2.15, and we have thus characterized all one-dimensional SSGM processes.

Remark 2.6.2. *As for Volterra representations in dimension one (cf. Section 2.4), note that the borderline case $H + M = 0$ leads to the degenerate process $X_t = at^H W(1)$, with $R(s,t) = a^2(st)^H$, which does not have a Volterra representation by Theorem 2.5.2. If $H + M < 0$, which is (2.4.1) for $d = 1$, Theorem 2.4.3 shows that the processes from Theorem 2.6.1 have a Volterra representation with kernels*

$$K(s,t) = \tilde{a} t^{H-1/2}(t/s)^{M+H+1/2}, \quad 0 < s < t,$$

where $\tilde{a} = a\sqrt{-2(H+M)}$.

2.7. Proving non-Markovianity of one-dimensional processes

As hinted at in the introduction, the initial goal was a unification and simplification of non-Markovianity proofs found in the literature. The following corollary should be applicable to virtually any self-similar non-Markovian Gaussian process with explicit covariance function. For example, it easily shows that the processes defined in [18, 65, 66, 67, 78] are not Markovian. Note that in [18] and [65] it is just stated that the standard condition (2.2.2) for Markovianity of Gaussian processes is easy to refute for the processes they consider. In [66, 67], the Markov property is not mentioned. In [78], (2.2.2) is used directly to prove non-Markovianity.

Definition 2.7.1. *Two real functions f and g are asymptotically equivalent at infinity, denoted by $f \sim g$, if $\lim_{x\to\infty} f(x)/g(x) = 1$.*

Corollary 2.7.2. *Let X be a centered self-similar one-dimensional Gaussian process with covariance $R(s,t) = \mathbb{E}[X_s X_t]$ and $R(1,1) = 1$. If the covariance $R(t^{1/2}, t)$ has an* *asymptotic expansion at infinity which contains more than one power of t, or a term that is not asymptotically equivalent to a power of t with coefficient 1, then X is not Markovian.*

Proof. If X is Markovian, then its covariance function satisfies $R(t^{1/2}, t) = t^{2H+M/2}$, by Theorem 2.6.1. This implies the assertion, by uniqeness of asymptotic expansions (see Section 1.5 in [28], or any other treatise on asymptotic expansions). $\square$

Other functions of t than $t^{1/2}$ could be used, of course, but in all examples from the literature we are aware of, putting $s = t^{1/2}$ seems convenient.

Example 1. The Riemann-Liouville process $X_t = \sqrt{2H} \int_0^t (t-s)^{H-1/2} dW_s$, where W is a Brownian motion and $H > 0$, is H-self-similar. We have $R(s,t) = s^{2H} l((t-s)/s)$ for $0 < s \le t$, where (see, e.g., [69])

$$l(u) := 2H \int_0^1 \big((v+u)v\big)^{H-1/2} dv = \frac{4Hu^{H-1/2}}{2H+1} {}_2F_1\left(\begin{matrix} \frac12 - H, H + \frac12 \\ H + \frac32 \end{matrix} \Big| -\frac{1}{u}\right).$$

Since the hypergeometric function ${}_2F_1$ is analytic at zero and ${}_2F_1(0) = 1$, we have

$$R(t^{1/2}, t) \sim \frac{4H}{2H+1} t^{3H/2-1/4}, \quad t \uparrow \infty.$$

By Corollary 2.7.2, the process can only be Markovian if $4H/(2H+1) = 1$, i.e. $H = \frac12$. We conclude that the Riemann-Liouville process is not Markovian for $H \in (0,\infty) \setminus \{\frac12\}$. While this has been mentioned in the literature, we are not aware of any published proof. Consequently, the volatility process in the rough Bergomi model [11] from financial mathematics, which is defined as an exponential of the Riemann-Liouville process, is not Markovian.

The generalized fractional Brownian motion of Pang and Taqqu [47, 61] is a centered Gaussian process with two parameters $\gamma \in [0,1)$ and $\alpha \in (-\frac12 - \frac12\gamma, \frac12 + \frac12\gamma)$. Its covariance function is

$$\begin{aligned} R(s,t) &= c^2 \int_0^s (t-u)^\alpha (s-u)^\alpha u^{-\gamma} du \\ &\quad + c^2 \int_0^\infty \big((t+u)^\alpha - u^\alpha\big)\big((s+u)^\alpha - u^\alpha\big) u^{-\gamma} du \\ &=: c^2 I_1(s,t) + c^2 I_2(s,t), \quad 0 \le s \le t, \end{aligned} \tag{2.7.1}$$

where $c > 0$ is a normalization constant that can be expressed in terms of the gamma and beta functions. This process is self-similar with exponent

$$H = \alpha - \frac{\gamma}{2} + \frac12.$$

The special case $\gamma = \alpha = 0$ is the standard Brownian motion. The Markov property has not been considered yet for generalized fractional Brownian motion.

Theorem 2.7.3. *Unless $\gamma = \alpha = 0$, the generalized fractional Brownian motion defined by the covariance function (2.7.1) is not a Markov process.*

Proof. We use Corollary 2.7.2, which requires the expansion of $R(\sqrt{t}, t)$ as $t \uparrow \infty$. The integral I_1 can be expressed by the hypergeometric function (see 15.6.1 in [30]),

$$I_1(\sqrt{t}, t) = \frac{\Gamma(\alpha+1)\Gamma(1-\gamma)}{\Gamma(\alpha-\gamma+2)} t^{3\alpha/2-\gamma/2+1/2} {}_2F_1\left(\begin{matrix} -\alpha, 1-\gamma \\ 2+\alpha-\gamma \end{matrix} \middle| \frac{1}{\sqrt{t}} \right). \tag{2.7.2}$$

In the integral I_2, we substitute $u = tv$ and split the integration range,

$$\begin{aligned} I_2(\sqrt{t}, t) &= t^{2\alpha-\gamma+1} \Big(\int_0^{t^{-1/4}} + \int_{t^{-1/4}}^{\infty} \Big) \big((1+v)^\alpha - v^\alpha \big) \big((t^{-1/2}+v)^\alpha - v^\alpha \big) v^{-\gamma} dv \\ &=: t^{2\alpha-\gamma+1} \big(I_{21}(\sqrt{t}, t) + I_{22}(\sqrt{t}, t) \big). \end{aligned} \tag{2.7.3}$$

For $\alpha \geq 0$ and $0 \leq v \leq t^{-1/4}$, we have $(1+v)^\alpha - v^\alpha \sim 1$, and thus

$$I_{21}(\sqrt{t}, t) \sim \int_0^{t^{-1/4}} \big((t^{-1/2}+v)^\alpha - v^\alpha \big) v^{-\gamma} dv,$$

which yields

$$I_{21}(\sqrt{t}, t) \sim \frac{1}{1-\gamma} t^{-\alpha/2+\gamma/4-1/4} {}_2F_1\left(\begin{matrix} -\alpha, 1-\gamma \\ 2-\gamma \end{matrix} \middle| -t^{1/4} \right) - \frac{t^{-\alpha/4+\gamma/4-1/4}}{\alpha-\gamma+1}. \tag{2.7.4}$$

If $\alpha < 0$, then $(1+v)^\alpha - v^\alpha \sim -v^\alpha$, and we obtain

$$\begin{aligned} I_{21}(\sqrt{t}, t) &\sim \int_0^{t^{-1/4}} \big(v^\alpha - (t^{-1/2}+v)^\alpha \big) v^{\alpha-\gamma} dv \\ &= \frac{t^{-\alpha/2+\gamma/4-1/4}}{2\alpha-\gamma+1} - \frac{t^{-3\alpha/4+\gamma/4-1/4}}{\alpha-\gamma+1} {}_2F_1\left(\begin{matrix} -\alpha, \alpha-\gamma+1 \\ \alpha-\gamma+2 \end{matrix} \middle| -t^{1/4} \right). \end{aligned}$$

For $t^{-1/4} < v$, we have $v\sqrt{t} \uparrow \infty$, hence

$$(t^{-1/2}+v)^\alpha - v^\alpha = \frac{\alpha}{\sqrt{t}} v^{\alpha-1} + \frac{\alpha(\alpha-1)v^{\alpha-2}}{2t} + O(v^{\alpha-3}t^{-3/2}), \tag{2.7.5}$$

and thus

$$I_{22}(\sqrt{t}, t) \sim \frac{\alpha}{\sqrt{t}} \int_{t^{-1/4}}^{\infty} \big((1+v)^\alpha - v^\alpha \big) v^{\alpha-\gamma-1} dv. \tag{2.7.6}$$

If $\alpha > \gamma$, this is $\alpha c_0 t^{-1/2} + O(t^{-3/4})$, because the integral

$$c_0 := \int_0^{\infty} \big((1+v)^\alpha - v^\alpha \big) v^{\alpha-\gamma-1} dv$$

exists. If $\alpha \le \gamma$, then (2.7.6) implies

$$I_{22}(\sqrt{t},t) \sim \frac{\alpha}{\sqrt{t}} \int_{t^{-1/4}}^{1/2} \big(1 + O(v) - v^{\alpha}\big) v^{\alpha-\gamma-1} dv \tag{2.7.7}$$

$$\sim \frac{\alpha}{\sqrt{t}} \int_{t^{-1/4}}^{1/2} (1 - v^{\alpha}) v^{\alpha-\gamma-1} dv \tag{2.7.8}$$

$$= \begin{cases} \frac{\alpha}{\sqrt{t}}\Big(\frac{t^{-\alpha/2+\gamma/4}}{2\alpha-\gamma} + \frac{t^{-\alpha/4+\gamma/4}}{\gamma-\alpha} + O(1)\Big), & \alpha < \gamma, \\ \frac{\gamma \log t + O(1)}{4\sqrt{t}}, & \alpha = \gamma. \end{cases}$$

From these estimates and the well-known expansion of the hypergeometric function, it is now a straightforward matter to check that the expansion of $R(\sqrt{t},t)$ as $t \uparrow \infty$ has at least two terms. A case distinction according to the values of α and γ is required, but as all cases are easy and rather similar, we just make the case where $\alpha > 0$ and $\alpha > \gamma$ explicit, and make this assumption throughout the rest of the proof. By (2.7.2) and §15.12(i) of [30],

$$I_1(\sqrt{t},t) = \frac{\Gamma(\alpha+1)\Gamma(1-\gamma)}{\Gamma(\alpha-\gamma+2)} t^{3\alpha/2-\gamma/2+1/2} + \frac{\alpha(\gamma-1)\Gamma(\alpha+1)\Gamma(1-\gamma)}{\Gamma(\alpha-\gamma+3)} t^{3\alpha/2-\gamma/2}\big(1+o(1)\big). \tag{2.7.9}$$

Using §15.12(i) of [30] in (2.7.4) shows that the first term of the expansion of ${}_2F_1$ cancels with the power of t at the end of (2.7.4), and the second term of the expansion of ${}_2F_1$ yields

$$I_{21}(\sqrt{t},t) \sim \frac{\alpha}{\alpha-\gamma} t^{-\alpha/4+\gamma/4-1/2}.$$

As noted after (2.7.6), $I_{22}(\sqrt{t},t) = \alpha c_0 t^{-1/2} + O(t^{-3/4})$. (Reasoning analogously to (2.7.8), it is easy to see that the term resulting from the second term on the right hand side of (2.7.5) is of lower order.) Since

$$-\frac{\alpha}{4} + \frac{\gamma}{4} - \frac{1}{2} > -\frac{3}{4},$$

we have

$$I_{21}(\sqrt{t},t) + I_{22}(\sqrt{t},t) = \alpha c_0 t^{-1/2} + \frac{\alpha}{\alpha-\gamma} t^{-\alpha/4+\gamma/4-1/2}\big(1+o(1)\big). \tag{2.7.10}$$

Then, by

$$2\alpha - \gamma + 1 + \Big(-\frac{\alpha}{4} + \frac{\gamma}{4} - \frac{1}{2}\Big) > \frac{3\alpha}{2} - \frac{\gamma}{2} + \frac{1}{2},$$

we see from (2.7.3), (2.7.9) and (2.7.10) that I_1 does not contribute to the first two terms of the expansion of (2.7.1), and so

$$R(\sqrt{t}, t) = \alpha c_0 c^2 t^{2\alpha-\gamma+1/2} + \frac{\alpha c^2}{\alpha - \gamma} t^{7\alpha/4-3\gamma/4+1/2}\big(1 + o(1)\big), \quad t \uparrow \infty.$$

As both coefficients in this expansion are clearly non-zero, and $\alpha > \gamma$ implies that the two exponents are different, we can apply Corollary 2.7.2. □

3. A characterization of real matrix semigroups

We characterize all real matrix semigroups, indexed by the non-negative reals, which satify a mild boundedness assumption, without assuming continuity. Besides the continuous solutions of the semigroup functional equation, we give a description of solutions arising from non-measurable solutions of Cauchy's functional equation. To do so, we discuss the primary decomposition and the Jordan–Chevalley decomposition of a matrix semigroup. Our motivation stems from a characterization of all multi-dimensional self-similar Gaussian Markov processes, which is given in a companion chapter 2.

3.1. Introduction

It is a classical fact that all continuous matrix-valued functions $g : [0, \infty) \to \mathbb{R}^{n\times n}$ satisfying the semigroup property

$$g(x + y) = g(x)g(y), \quad x, y \geq 0, \tag{3.1.1}$$

$$g(0) = \mathrm{id}, \tag{3.1.2}$$

where id is the identity matrix, are given by the maps $g(x) = \exp(Mx)$, where $M \in \mathbb{R}^{n\times n}$ is arbitrary. See, e.g., Problem 2.1 in [35] and the subsequent discussion. To the best of our knowledge, few authors have considered non-continuous solutions of (3.1.1) in the multidimensional case $d > 1$. Kuczma and Zajtz [51] determine all matrix semigroups $(g(x))_{x\in\mathbb{R}}$ where $x \mapsto g(x)$ is measurable. Zajtz [75] characterizes the matrix semigroups $(g(x))_{x\in\mathbb{Q}}$ indexed by rational numbers. The one-dimensional case, which is equivalent to Cauchy's functional equation,

$$f(x) + f(y) = f(x + y), \quad f : \mathbb{R} \to \mathbb{R}, \tag{3.1.3}$$

is well studied, on the other hand (see [4, 16]). Our motivation to investigate non-continuous matrix semigroups stems from probability theory: In chapter 2, we develop a characterization of all self-similar Gaussian Markov processes. By self-similarity, the bivariate covariance function of such a multi-dimensional stochastic process can be transformed to a matrix function of a single argument, which must satisfy (3.1.1). See Section 3.2 for some more details. Our main result (Theorem 3.2.6) determines all solutions of (3.1.1) which satisfy a mild boundedness assumption. In Section 3.3, we prepare the proof by providing a decomposition of the vector space into invariant subspaces, and establish some useful properties of the decomposition. The main proofs (in Section 3.5) are preceded by a discussion of the Jordan–Chevalley decomposition of a semigroup, in Section 3.4. In Appendix 3.A, we give two auxiliary results on Cauchy's functional equation.

3.2. Preliminaries and main result

To state our main result, define the rotation matrix

$$Q(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}, \quad \theta \in \mathbb{R},$$

and, for even k and a function $\nu : \mathbb{R} \to \mathbb{R}$, the block-diagonal matrix

$$Q_k^\nu(x) := \begin{pmatrix} Q(\nu(x)) & & 0 \\ & \ddots & \\ 0 & & Q(\nu(x)) \end{pmatrix} \in \mathbb{R}^{k\times k} \tag{3.2.1}$$

consisting of $k/2$ rotation matrices. In our statements, ν will denote some non-measurable (equivalently, non-continuous) solution of (3.1.3). The following assumption is in force throughout the chapter.

Assumption 3.2.1. *In the following, V denotes a real d-dimensional vector space equipped with an inner product $\langle \cdot, \cdot \rangle$.*

Definition 3.2.2. *We write $L(V)$ for the set of linear maps from V to itself. The operator norm induced by the inner product $\langle \cdot, \cdot \rangle$ is denoted by $\| \cdot \|_{\mathrm{op}}$. If the basis is clear from the context, we will identify elements of $L(V)$ and $d \times d$ matrices.*

Definition 3.2.3. *We say that $g : \mathbb{R}_{\geq 0} \to L(V)$ is a semigroup if $g(0) = \mathrm{id}$ and*

$$g(x + y) = g(x)g(y)$$

for all $x, y \geq 0$. Semigroups acting on $V^{\mathbb{C}}$, the complexification of V, are defined by the same property.

Definition 3.2.4. *We say that a semigroup $(g(x))_{x\geq 0}$ in $L(V)$ is elementary if there exists an orthonormal basis such that*

$$g(x) = \exp(Mx),\ x \geq 0, \quad \text{or} \quad g(x) = Q_d^\nu(x) \exp(Mx),\ x \geq 0,$$

for some non-continuous ν satisfying Cauchy's equation (3.1.3) and some matrix $M \in \mathbb{R}^{d\times d}$.

Assumption 3.2.5. *The semigroup $g(x)_{x\geq 0}$ satisfies $\|g(x)\|_{\mathrm{op}} \leq f(x)$ for $x \geq 0$, where the function $f : [0, \infty) \to \mathbb{R}$ is locally bounded, right-continuous at 0 and satisfies $f(0) = 1$.*

We can now state our main theorem, which will be proven at the end of Section 3.5.

Theorem 3.2.6. *Let $(g(x))_{x\geq 0}$ be a semigroup satisfying Assumption 3.2.5. Then there exists an orthogonal decomposition $V = \bigoplus_{i=1}^{n} V_i$ such that each V_i is invariant under $g(x)$, and either $g(x)$ is elementary on V_i or $g(x)|_{V_i} = 0$ for $x > 0$.*

We call $g((x))_{x\geq 0}$ degenerate if $g(x) = 0$ for $x > 0$.

Example 2. Let $c \in \mathbb{R}$, and let $f : \mathbb{R} \to \mathbb{R}$ be a non-continuous solution of Cauchy's functional equation (3.1.3). Then the semigroup given by

$$g(x) := \begin{pmatrix} \cos(f(x)) & -\sin(f(x)) \\ \sin(f(x)) & \cos(f(x)) \end{pmatrix} e^{cx}, \quad x \geq 0,$$

is an example of a semigroup covered by Theorem 3.2.6 (by putting $d = 2$ and $M = c\,\mathrm{id}$ in Definition 3.2.4), but not the previous results mentioned at the beginning of the introduction. It illustrates that, in contrast to dimension one, a two-dimensional locally bounded semigroup need not be continuous.

3.3. Primary decomposition for semigroups

In this section, we discuss a decomposition of V into subspaces which are invariant for the given semigroup.

Definition 3.3.1. *Let S be a linear operator on the vector space V. A subspace $U \subseteq V$ is called S-invariant if S maps U into U.*

The primary decomposition theorem from linear algebra [59, Theorem 1.5.1] decomposes a vector space into invariant subspaces for a given operator. On each subspace, the operator has a single real eigenvalue or a pair of conjugate complex eigenvalues. Instead of a single operator, we need the following version for semigroups.

Theorem 3.3.2 (Primary decomposition)**.** *For any semigroup $(g(x))_{x\geq 0}$ of linear maps acting on V there exists a decomposition $V = \bigoplus_{i=1}^{n} V_i$ with $\dim(V_i) \geq 1$ such that each V_i is $g(x)$-invariant for all $x \geq 0$, and for all i one of the following holds:*

(1) For all $x \geq 0$, $g(x)$ has one eigenvalue $\lambda(x) \geq 0$ on V_i,

(2) For all $x \geq 0$, $g(x)$ has eigenvalues $\lambda(x), \overline{\lambda}(x) \in \mathbb{C}$ on V_i. They may coincide (and thus be real) for some values of x, but not for all $x \geq 0$.

Proof. It is known that the primary decomposition extends to commuting sets of matrices. Indeed, the decomposition $V = \bigoplus_{i=1}^{n} V_i$ into invariant subspaces follows from Theorem 5 on p. 40 in [48], applied to the span of the semigroup $(g(x))_{x\geq 0}$. Thus, we only need to argue why (1) or (2) follows from the semigroup property. Assume that each $g(x)$ has only one eigenvalue $\lambda(x) \in \mathbb{R}$ on some V_i. Since the $g(x)$ commute, they share a common eigenvector v_i, and thus we have $\lambda(x)\lambda(y) = \lambda(x + y)$ for $x, y \geq 0$. Suppose $\lambda(x_0) < 0$ for some $x_0 > 0$. Then we have $\lambda(x_0) = \lambda(x_0/2)^2$, and hence $\lambda(x_0/2) \notin \mathbb{R}$, contradicting $\lambda(x) \in \mathbb{R}$ for all $x \geq 0$. Hence $\lambda(x) \in \mathbb{R}_{\geq 0}$ for all $x \geq 0$. $\square$

Theorem 3.3.3. *Consider a semigroup $g = (g(x))_{\geq 0}$ acting on $V^{\mathbb{C}}$, the complexification of V. Then there exists a decomposition $V^{\mathbb{C}} = \bigoplus_{i=1}^{n} V_i$ with $\dim(V_i) \geq 1$ such that, for each i and all $x \geq 0$, the space V_i is $g(x)$-invariant and $g(x)$ has only one eigenvalue $\lambda(x) \in \mathbb{C}$ on V_i.*

Proof. This is an immediate consequence of Theorem 5 on p. 40 in [48] (cf. the preceding proof). □

Definition 3.3.4. *We call the decomposition from Theorem 3.3.2 simultaneous real primary decomposition (SRPD) of V, omitting "w.r.t. g" if the semigroup is clear from the context. The component V_i is of first type in case (1), and of second type in case (2). Similarly, the simultaneous primary decomposition (SPD) is the decomposition from Theorem 3.3.3.*

Lemma 3.3.5. *Let g be a semigroup acting on V, and let $V_i \subseteq V$ be a subspace of first type from the SRPD of V. Then there exists a common eigenvector $v \in V$ such that $g(x)v = \lambda(x)v$ for all $x > 0$.*

Proof. We present an algorithm which yields the subspace of common eigenvectors. If $V_i =: V_i^{(0)}$ itself is this subspace, then we are done. Otherwise there exists $x_0 > 0$ such that the eigenspace $V_i^{(1)}$ of $g(x_0)$ is a strict subspace of V_i. For any $y > 0$ and any $v \in V_i^{(1)}$ we have

$$\begin{aligned} 0 = g(y)0 &= g(y)[\lambda(x_0)v - g(x_0)v] \\ &= \lambda(x_0)[g(y)v] - g(x_0)[g(y)v]. \end{aligned}$$

It follows that $V_i^{(1)}$ is $g(x)$-invariant for all $x > 0$. Now either $V_i^{(1)}$ consists of common eigenvectors, or we can again find $x_1 > 0$ such that the eigenspace $V_i^{(2)}$ of $g(x_1)$ in $V_i^{(1)}$ is a strict subspace of $V_i^{(1)}$. Repeating this argument yields a sequence of nontrivial subspaces $V_i^{(n)}$ whose dimensions are strictly decreasing, hence it has to terminate. Clearly, the final vector space in this sequence is the space of common eigenvectors of the semigroup. □

Lemma 3.3.6. *Let g be a semigroup acting on V, and let V_i be a subspace from the SRPD of V such that there exists $x_0 > 0$ with $\lambda(x_0) = 0$. Then we have $\lambda(x) = 0$ for all $x > 0$. Furthermore $V_i \subseteq \ker(g(x))$ for all $x > 0$.*

Proof. By the previous lemma there exists a common eigenvector $v \in V_i$. (Since commuting matrices are simultaneously triangularizable it follows that they share a common eigenvector $v \in V^{\mathbb{C}}$.) We have

$$\lambda(x + y)v = g(x + y)v = g(x)g(y)v = \lambda(x)\lambda(y)v.$$

Hence λ satisfies $\lambda(x)\lambda(y) = \lambda(x+y)$. Since $\lambda(x_0) = 0$ we have $\lambda(x_0/n) = 0$ for all $n \in \mathbb{N}$ since $\lambda(x_0/n)^n = 0$. Let $y > 0$, then there exists $n \in \mathbb{N}$ such that $x_0/n < y$. We obtain

$$\lambda(y) = \lambda(x_0/n)\lambda(y - x_0/n) = 0.$$

Since $\lambda(y) = 0$ for all $y > 0$ it follows that the characteristic polynomial of $g(y)$ satisfies $\chi_{g(y)}(Z) = Z^{\dim(V_i)}$ and hence, by the Cayley–Hamilton theorem, $g(y)^{\dim(V_i)} \equiv 0$. For $v \in V_i$ we obtain

$$g(y)v = g(y/\dim(V_i))^{\dim(V_i)}v = 0. \qquad \square$$

Corollary 3.3.7. *Let g be a semigroup acting on V, with SRPD $V = \bigoplus_{i=1}^n V_i$. For any $x > 0$ we have $\ker(g(x)) = \bigoplus_{i\in I} V_i$, where*

$$I = \{i \,|\, V_i \text{ of type 1 with } \lambda \equiv 0\}.$$

Proof. If $V_i \subseteq \ker(g(x))$, then $0 = \lambda(x) \in \mathbb{R}$ for all $x \geq 0$ and hence V_i is of type 1. By the previous lemma we have

$$\ker(g(x)) \supseteq \bigoplus_{i\in I} V_i.$$

If the inclusion was strict, then there would exist V_j with $j \notin I$ such that $\ker(g(x)) \cap V_j \neq \emptyset$. But since $g(x)$ is invertible on V_j this gives a contradiction, hence we have equality. $\square$

Corollary 3.3.8. *Let g be a semigroup acting on V. Then there exists a decomposition $V = V_1 \oplus V_2$, such that $g(x)|_{V_1}$ is invertible for all $x \geq 0$ and $g(x)|_{V_2} \equiv 0$ for $x > 0$.*

3.4. Multiplicative Jordan–Chevalley decomposition of semigroups

Due to Corollary 3.3.8, from now on we assume in most of our statements that $g(x)$ is invertible for all $x \geq 0$. A standard result from linear algebra, the Jordan–Chevalley decomposition, asserts that any matrix A can be uniquely decomposed as $A = D + N$, where D is diagonalizable, N is nilpotent and D and N commute. If A is invertible, then we can express it as $A = D(\mathrm{id} + D^{-1}N) := DT$ with T unipotent and commuting with D.

Definition 3.4.1. *For an invertible linear map A on $\mathbb{K}^d$ with $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, the multiplicative decomposition $A = DT$ into commuting factors with D diagonalizable and T unipotent, is called the multiplicative Jordan–Chevalley decomposition.*

For background on the (multiplicative) Jordan–Chevalley decomposition, we refer to Section 15.1 in [44]. We now analyze the structure of the multiplicative Jordan–Chevalley decomposition of a semigroup.

Theorem 3.4.2. *Let $(g(x))_{x\geq 0}$ be a semigroup of invertible linear maps acting on $\mathbb{K}^d$ with $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $g(x) = D(x)T(x)$ be the multiplicative Jordan–Chevalley decomposition of each $g(x)$. Then $((D(x))_{x\geq 0})$ and $(T(x))_{x\geq 0}$ each form a semigroup, and the two families commute with each other, i.e. $T(x)D(y) = D(y)T(x)$ for all $x, y \geq 0$.*

Proof. We can w.l.o.g. assume that $\mathbb{K} = \mathbb{C}$, since by uniqueness of the Jordan–Chevalley decomposition $D(x) \in \mathbb{R}^{d\times d}$ and $T(x) \in \mathbb{R}^{d\times d}$ if $g(x) \in \mathbb{R}^{d\times d}$ for $x \geq 0$. Take the SPD from Theorem 3.3.3, $V = \bigoplus_{i=1}^{k} V_i$, so that each $g(x)$ has only one eigenvalue on V_i. Denote by $g(x)|_{V_i} = D_i(x)T_i(x)$ the multiplicative Jordan–Chevalley decomposition $g(x)$ restricted to V_i. Denote by $\lambda_i(x)$ the eigenvalue of $g(x)$ on V_i. Clearly, $D_i(x) = \lambda_i(x)\,\mathrm{id}$. Since the $(g_i(x))_{x\geq 0}$ are a commuting family of matrices, they share a common eigenvector $v_i \in V_i$. We have

$$\lambda_i(x)\lambda_i(y)v_i = g_i(x)g_i(y)v_i = g_i(x+y)v_i = \lambda_i(x+y)v_i,$$

and hence $(D_i(x))_{x\geq 0}$ is a semigroup. Since each $D_i(x)$ is a multiple of the identity, it commutes with every linear map and hence

$$T_i(x)T_i(y) = \frac{1}{\lambda_i(x)\lambda_i(y)}g_i(x)g_i(y) = \frac{1}{\lambda_i(x+y)}g_i(x+y) = T_i(x+y),$$

which shows that $(T_i(x))_{x\geq 0}$ is also a semigroup. The result then follows, since by uniqueness $(\bigoplus D_i(x)) \otimes (\bigoplus T_i(x))$ is the multiplicative Jordan–Chevalley decomposition of $g(x)$. $\square$

Theorem 3.4.3. *Let $(g(x))_{x\geq 0}$ be a semigroup of invertible linear maps acting on V and let $g(x) = D(x)T(x)$ be its multiplicative Jordan–Chevalley decomposition. Then there exist commuting real diagonalizable linear maps $J(x)$ and commuting real nilpotent linear maps $N(x)$ satisfying*

1. $J(x) + J(y) = J(x+y)$
2. $N(x) + N(y) = N(x+y)$
3. $J(x)N(y) = N(y)J(x)$ *for all* $x, y \geq 0$

such that

$$D(x) = \exp(J(x)), \quad T(x) = \exp(N(x))$$

and

$$g(x) = \exp(J(x) + N(x)).$$

Proof. Let $V = \bigoplus_{i=1}^{n} V_i$ be the SRPD and let $g_i(x) := g(x)|_{V_i}$. Assume first that V_i is of first type and $\lambda_i(x)$ is the single positive eigenvalue of $g_i(x)$. Then the multiplicative Jordan–Chevalley decomposition on V_i is $g_i(x) = \lambda_i(x)\,\mathrm{id}\,T_i(x)$. Define $J_i(x) := \log(\lambda_i(x))\,\mathrm{id}$. Since $T_i(x) - \mathrm{id}$ is nilpotent, we have

$$N_i(x) := \log(T_i(x)) = \sum_{k=1}^{d-1} \frac{1}{k}(\mathrm{id} - T_i(x))^k.$$

Notice that since the logarithm converges for all unipotent matrices, the exponential map between the Lie algebra of nilpotent matrices and the Lie group of unipotent matrices is bijective (see p. 35 in [41]). Since $T_i(x)$ is a semigroup we have

$$\exp(N_i(x))\exp(N_i(y)) = \exp(N_i(y))\exp(N_i(x)) = \exp(N_i(x+y)).$$

Rewrite this as

$$\exp(N_i(y))^{-1}\exp(N_i(x))\exp(N_i(y)) = \exp\big(\exp(N_i(y))^{-1}N_i(x)\exp(N_i(y))\big) = \exp(N_i(x)).$$

Since $\exp(N_i(y))^{-1}N_i(x)\exp(N_i(y))$ is nilpotent and exp is bijective, we have

$$\exp(N_i(y))^{-1}N_i(x)\exp(N_i(y)) = N_i(x).$$

For any $t \in \mathbb{R}$ set $N_i^t(x) := t\,\mathrm{id} + N_i(x)$, which is invertible for $t \neq 0$. It is clear that

$$N_i^t(x)\exp(N_i(y)) = \exp(N_i(y))N_i^t(x).$$

Applying the same idea again yields

$$\exp(N_i^t(x)^{-1}N_i(y)N_i^t(x)) = \exp(N_i(y)).$$

Again by uniqueness we obtain

$$N_i^t(x)N_i(y) = N_i(y)N_i^t(x)$$

or equivalently

$$N_i(x)N_i(y) = N_i(y)N_i(x).$$

By commutativity we obtain

$$\exp(N_i(x) + N_i(y)) = \exp(N_i(x))\exp(N_i(y)) = \exp(N_i(x+y))$$

and hence, again by uniqueness, we have $N_i(x) + N_i(y) = N_i(x+y)$. It follows that $J_i(x)$ and $N_i(x)$ have the desired properties.

In the second case V_i is of second type. By Theorem 3.3.3, and since $g_i(x)$ is real, V_i decomposes over $\mathbb{C}$ as $V_i = U_1 \oplus U_2$, where $g_i(x)$ has only one eigenvalue on each U_j and U_1 is isomorphic to U_2 with isomorphism given by $u \in U_1 \mapsto \overline{u} \in U_2$. By taking the principal branch of the logarithm, in the same manner as in the real case we obtain commuting J, N on U_1 and $\overline{J}, \overline{N}$ on U_2 satisfying Cauchy's equation such that

$$g_i(x) = \exp(J(x) + N(x)) \oplus \exp(\overline{J}(x) + \overline{N}(x)).$$

Notice that $J(x) = (\mu(x) + \mathbf{i}\nu(x))\,\mathrm{id}$, where $\mathbf{i} = \sqrt{-1}$ and $\nu(x)$ satisfies Cauchy's functional equation on $\mathbb{R}/[-\pi,\pi)$. By Lemma 3.A.1, we can lift any solution on $\mathbb{R}/[-\pi,\pi)$ to a solution on $\mathbb{R}$ such that linearity is preserved. From now on denote this lift by $\nu(x)$. Hence if we choose any basis of U_1 and its complex conjugate on U_2 we obtain that $g_i(x)$

is similar to

$$g_i(x) \sim \exp\left[\begin{pmatrix} J(x)+N(x) & 0 \\ 0 & \overline{J}(x)+\overline{N}(x) \end{pmatrix}\right].$$

Taking the similarity transform with the matrix

$$A := \frac{1}{\sqrt{2}}\begin{pmatrix} \mathrm{id} & \mathbf{i}\cdot\mathrm{id} \\ \mathbf{i}\cdot\mathrm{id} & \mathrm{id} \end{pmatrix},$$

where $\mathrm{id} = \mathrm{id}_{\dim(U_1)}$, we obtain

$$\begin{aligned} &A\begin{pmatrix} J(x)+N(x) & 0 \\ 0 & \overline{J}(x)+\overline{N}(x) \end{pmatrix}A^{-1} \\ &\qquad\qquad = \begin{pmatrix} \mathrm{Re}(J(x)+N(x)) & \mathrm{Im}(J(x)+N(x)) \\ -\mathrm{Im}(J(x)+N(x)) & \mathrm{Re}(J(x)+N(x)) \end{pmatrix} \in L(V_i). \end{aligned}$$

Since matrix similarity over $\mathbb{C}$ is equivalent to matrix similarity over $\mathbb{R}$ for two real matrices, there exists a real matrix B_i on V_i such that

$$g_i(x) = \exp\left[B_i\begin{pmatrix} \mathrm{Re}(J(x)) & \mathrm{Im}(J(x)) \\ -\mathrm{Im}(J(x)) & \mathrm{Re}(J(x)) \end{pmatrix}B_i^{-1} + B_i\begin{pmatrix} \mathrm{Re}(N(x)) & \mathrm{Im}(N(x)) \\ -\mathrm{Im}(N(x)) & \mathrm{Re}(N(x)) \end{pmatrix}B_i^{-1}\right]. \tag{3.4.1}$$

Setting

$$J_i(x) := B_i\begin{pmatrix} \mathrm{Re}(J(x)) & \mathrm{Im}(J(x)) \\ -\mathrm{Im}(J(x)) & \mathrm{Re}(J(x)) \end{pmatrix}B_i^{-1} = B_i\begin{pmatrix} \mathrm{id}\,\mu(x) & \mathrm{id}\,\nu(x) \\ -\mathrm{id}\,\nu(x) & \mathrm{id}\,\mu(x) \end{pmatrix}B_i^{-1}$$

and

$$N_i(x) := B_i\begin{pmatrix} \mathrm{Re}(N(x)) & \mathrm{Im}(N(x)) \\ -\mathrm{Im}(N(x)) & \mathrm{Re}(N(x)) \end{pmatrix}B_i^{-1},$$

we see that J_i and N_i satisfy the desired conditions, with $D_i(x) = \exp(J_i(x))$ and $T_i(x) = \exp(N_i(x))$ by uniqueness of the Jordan–Chevalley decomposition. The direct sums $\bigoplus_{i=1}^n J_i(x)$ and $\bigoplus_{i=1}^n N_i(x)$ give the required matrices. Furthermore in the case where there exists $x > 0$ for which $g_i(x)$ has a complex eigenvalue $\lambda(x) = \mu(x) + \mathbf{i}\nu(x)$ we have

$$g_i(x) = \underbrace{B_i\begin{pmatrix} \cos(\nu(x))\,\mathrm{id} & \sin(\nu(x))\,\mathrm{id} \\ -\sin(\nu(x))\,\mathrm{id} & \cos(\nu(x))\,\mathrm{id} \end{pmatrix}B_i^{-1}}_{U_i(x)}\exp(\mu(x)\,\mathrm{id} + B_iN(x)B_i^{-1}),$$

where $U_i(x) \in SO(\dim(V_i))$. Recall the matrix defined in (3.2.1). By changing the order of the basis, we have that U_i is similar to the block diagonal matrix (recall the notation (3.2.1))

$$U_i(x) \sim \begin{pmatrix} Q(\nu(x)) & & \\ & \ddots & \\ & & Q(\nu(x)) \end{pmatrix} = Q^{\nu}_{\dim(V_i)}(x). \tag{3.4.2}$$

Hence

$$g_i(x) = \tilde{B}_i Q^{\nu}_{\dim(V_i)}(x) \tilde{B}_i^{-1} \exp(\mu(x)\,\mathrm{id} + B_i N(x) B_i^{-1}), \tag{3.4.3}$$

where $\tilde{B}_i$ is the composition of B_i with some permutation matrix P. □

3.5. Proof of the main result

After providing some final preparatory results, this section ends with the proof of Theorem 3.2.6. Consider again the SRPD from Theorem 3.3.2. Now on each V_i, $g(x)$ has either one positive eigenvalue $\lambda_i(x) = e^{\mu_i(x)}$ or two complex conjugate eigenvalues $\lambda_i(x) = e^{\mu_i(x)+\mathbf{i}\nu_i(x)}$ and $\overline{\lambda_i}(x) = e^{\mu_i(x)-\mathbf{i}\nu_i(x)}$, where it is possible that $\nu_i(x) = \pi$ for some values of x. If V_i is of first type, set $\nu_i(x) = 0$. Each ν_i is a solution to Cauchy's functional equation (3.1.3). Consider then the set $\{f | f = \nu_i \text{ or } f = -\nu_i \text{ for some } 1 \leq i \leq n\}$ and partition it into equivalent solutions, according to Definition 3.A.2. This then gives a partition of the index set $\{1, \ldots, n\} = \bigcup_{l=1}^k I_l$ in the following manner: If $\nu_i \sim \nu_j$ or $\nu_i \sim -\nu_j$, then i, j are in the same subset of the partition. This is well-defined, since if $f \sim g$ then $-f \sim -g$. Set $W_l := \bigoplus_{i \in I_l} V_i$.

Definition 3.5.1. *We call the decomposition $V = \bigoplus_{i=1}^k W_i$ the partitioned SRPD of V.*

Furthermore associate with each W_j one solution η_j of Cauchy's equation such that $\eta_j \sim \nu_i$ with $i \in I_j$. If η_j is linear we always take $\eta_j \equiv 0$. Notice that for $i \neq j$ we have $\eta_i \not\sim \eta_j$, hence there can be at most one W_i with $\eta_i = 0$, and furthermore this is the only W_i which can have odd dimension since it contains all V_j of type 1 (recall Definition 3.3.4).

Theorem 3.5.2. *Let $(g(x))_{x \geq 0}$ be a semigroup of invertible linear maps acting on V, and denote by $V = \bigoplus_{i=1}^k W_i$ its partitioned SRPD. Denote by η_i the solution associated with W_i. If η_i is non-continuous, then there exists a change of basis A_i on W_i such that*

$$g(x)|_{W_i} = A_i Q^{\eta_i}_{d_i}(x) A_i^{-1} \exp(G_i(x)),$$

where $d_i = \dim(W_i)$, and

$$A_i Q^{\eta_i}_{d_i}(x) A_i^{-1} G_i(y) = G_i(y) A_i Q^{\nu_i}_{d_i}(x) A_i^{-1}$$

for all $x, y \geq 0$. If $\eta_i = 0$, then

$$g(x)|_{W_i} = \exp(G_i(x)).$$

In both cases $G_i(x) : W_i \to W_i$ is a commuting family of matrices on W_i satisfying Cauchy's functional equation. Furthermore, if v is a common eigenvector of $(G_i(x))_{x\geq 0}$ such that $G_i(x)v = [\mu(x) + \mathbf{i}\nu(x)]v$, then ν is linear in x.

Proof. Assume first that η_i, the associated solution of Cauchy's functional equation, is non-continuous. For each W_i we have the decomposition $W_i = \bigoplus_{j\in I_i} V_j$ where V_j are subspaces from the SRPD. Since η_i is non-continuous, each V_j in the direct sum has to be of second type, hence by (3.4.3) on each V_j we have

$$g_j(x) = \tilde{B}_j Q^{\nu_j}_{\dim(V_j)}(x)\tilde{B}_j^{-1} \exp\big(\mu(x)\,\mathrm{id} + B_j N(x) B_j^{-1}\big).$$

By definition of η_i we have $\eta_i \sim \nu_j$ and hence there exists $c_j \in \mathbb{R}$ such that $\nu_j(x) = \eta_i(x) + c_j x$. Let

$$C_j := c_j \begin{pmatrix} L & & \\ & \ddots & \\ & & L \end{pmatrix}$$

with

$$L := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

where C_j is of dimension $\dim(V_j) \times \dim(V_j)$. Then we have

$$\begin{aligned} g_j(x) &= \tilde{B}_j Q^{\nu_j}_{\dim(V_j)}(x)\tilde{B}_j^{-1} \exp\big(\mu_j(x)\,\mathrm{id} + B_j N_j(x) B_j^{-1}\big) \\ &= \tilde{B}_j Q^{\eta_i}_{\dim(V_j)}(x)\exp(C_j x)\tilde{B}_j^{-1} \exp\big(\mu_j(x)\,\mathrm{id} + B_j N_j(x) B_j^{-1}\big) \\ &= \tilde{B}_j Q^{\eta_i}_{\dim(V_j)}(x)\tilde{B}_j^{-1} \exp\big(\tilde{B}_j C_j x \tilde{B}_j^{-1} + \mu_j(x)\,\mathrm{id} + B_j N_j(x) B_j^{-1}\big). \end{aligned}$$

Set $\tilde{G}_j(x) := \tilde{B}_j C_j x \tilde{B}_j^{-1} + \mu_j(x)\,\mathrm{id} + B_j N_j(x) B_j^{-1}$. Then

$$A_i := \bigoplus_{j\in I_i} \tilde{B}_j \quad \text{and} \quad G_i(x) := \bigoplus_{j\in I_i} \tilde{G}_j(x) \tag{3.5.1}$$

satisfy

$$g(x)|_{W_i} = \bigoplus_{j\in I_i} g_j(x) = A_i Q^{\eta_i}_{d_i}(x) A_i^{-1} \exp(G_i(x)).$$

By construction the common eigenvalues of $\exp(G_i(x))$ satisfy $\lambda(x) = e^{\mu_j(x) \pm \mathbf{i} c_j x}$. If $\eta_i \equiv 0$, then by Theorem 3.4.3 there exists $G_i(x)$ such that $g(x)|_{W_i} = \exp(G_i(x))$. By the definition of W_i, it follows that the imaginary parts of the eigenvalues of $G_i(x)$ have to be equivalent to $\eta_i \equiv 0$, hence they have to be linear. $\square$

Theorem 3.5.3. *Let g be a non-degenerate semigroup acting on V which satisfies Assumption 3.2.5. Then there exists a matrix M and a semigroup $S(x) \in SO(d)$, where*

$SO(d)$ is the set of special orthogonal matrices, such that

$$g(x) = S(x)\exp(Mx), \quad x \geq 0,$$

with M commuting with $S(x)$.

Proof. Consider the SRPD $V = \bigoplus_{i=1}^{n} V_i$ from Theorem 3.3.2, and define $g_i(x) := g(x)|_{V_i}$. Clearly, $\|g_i(x)\|_{\mathrm{op}} \leq \|g(x)\|_{\mathrm{op}} \leq f(x)$. For each eigenvalue $\lambda_i(x) = e^{\mu_i(x)+\mathbf{i}\nu_i(x)}$ on V_i we have

$$|\lambda_i(x)| \leq \|g_i(x)\|_{\mathrm{op}} \leq f(x).$$

Since $\lambda_i(x)\lambda_i(y) = \lambda_i(x+y)$, we have $|\lambda_i(x)||\lambda_i(y)| = |\lambda_i(x+y)|$. Moreover, $\mu_i(x) = \log(|\lambda_i(x)|)$, and so $\mu_i(x) \leq \log(f(x))$ for $x \geq 0$ and $\mu_i(x) + \mu_i(y) = \mu_i(x+y)$. It follows that μ_i is locally bounded and hence that $\mu_i(x) = a_i x$ for some $a_i \geq 0$. As in the proof of Theorem 3.4.3, on V_i we have $g_i(x) = \exp(J_i(x))T_i(x)$ where $J_i(x)$ is diagonalizable with eigenvalues $a_i x \pm \mathbf{i}\nu_i(x)$, and thus

$$\|T_i(x)\|_{\mathrm{op}} \leq f(x)e^{-a_i x}.$$

As in Theorem 3.4.3, set

$$N_i(x) := \log(T_i(x)) = \sum_{k=1}^{d-1} \frac{1}{k}(\mathrm{id} - T_i(x))^k.$$

Then, since $\|\mathrm{id} - T_i(x)\|_{\mathrm{op}} \leq 1 + f(x)e^{-a_i x}$, we obtain

$$\|N_i(x)\|_{\mathrm{op}} \leq \sum_{k=1}^{d-1} \frac{1}{k}\|\mathrm{id} - T_i(x)\|_{\mathrm{op}}^k \leq \sum_{k=1}^{d-1} \frac{1}{k}\left(1 + f(x)e^{-a_i x}\right)^k =: F(x), \tag{3.5.2}$$

where $F(x)$ is again locally bounded. Since the operator norm in finite dimensions is equivalent to the L^∞-norm, it follows that each entry of $N_i(x)$ is locally bounded and satisfies Cauchy's functional equation. Thus, there exists a nilpotent linear map P_i such that $N_i(x) = P_i x$. Let $V = \bigoplus_{i=1}^{k} W_i$ be the partitioned SRPD of V. Assume first that the solution associated with W_i is $\eta_i \equiv 0$. Since $W_i = \bigoplus_{j\in I_i} V_j$ we have $g(x)|_{W_i} = \bigoplus_{j\in I_i} \exp(J_j(x) + P_j x)$. The real part of the eigenvalues of $J_j(x)$ is linear in x and by Theorem 3.5.2 also the complex parts have to be linear since $\nu_i \equiv 0$. Since $J_j(x)$ is diagonalizable and all its eigenvalues are continuous in x, $J_j(x)$ is continuous in x and hence $J_j(x) = \tilde{M}_j x$. Setting $M_i := \bigoplus_{j\in I_i}(\tilde{M}_j + P_j)$ and $S_i(x) = \mathrm{id}$ yields $g(x)|_{W_i} = S_i(x)\exp(M_i x)$. By Theorem 3.5.2 and (3.5.1), we have for W_i with η_i non-continuous

$$g(x)|_{W_i} = A_i Q_{d_i}^{\eta_i}(x) A_i^{-1} \exp(G_i(x))$$

with

$$G_i(x) = \bigoplus_{j\in I_i} \big(\tilde{B}_j C_j x \tilde{B}_j^{-1} + \mu_j(x)\,\mathrm{id} + B_j N_j(x) B_j^{-1}\big)$$
$$= \bigoplus_{j\in I_i} \big(\tilde{B}_j C_j \tilde{B}_j^{-1} + a_j\,\mathrm{id} + B_j P_j B_j^{-1}\big) x.$$

Hence, setting $M_i := \bigoplus_{j\in I_i}(\tilde{B}_j C_j \tilde{B}_j^{-1} + a_j\,\mathrm{id} + B_j P_j B_j^{-1})$, we have $G_i(x) = M_i x$. Thus

$$g(x)|_{W_i} = A_i Q_{d_i}^{\eta_i}(x) A_i^{-1} \exp(M_i x).$$

Next we show that $S_i(x) := A_i Q_{d_i}^{\eta_i}(x) A_i^{-1}$ is an isometry on W_i. Since $\|g(x)\|_{\mathrm{op}} \leq f(x)$, we have

$$\langle v, v\rangle - \frac{1}{f(x)^2}\langle g(x)v, g(x)v\rangle \geq 0 \tag{3.5.3}$$

for any $v \in V$. Hence, for $v \in W_i$ we have

$$\langle v, v\rangle - \frac{1}{f(x)^2}\big\langle \exp(M_i x) S_i(x) v, \exp(M_i x) S_i(x) v\big\rangle \geq 0.$$

Fix an arbitrary $x_0 \geq 0$, and set $\theta_0 := \eta_i(x_0)$. The graph of η_i is dense in $\mathbb{R}\times\mathbb{R}$ (see Appendix 3.A), and so there is a sequence of positive reals x_n with $\lim_{n\to\infty} x_n = 0$ such that $\lim \eta_i(x_n) = \theta_0$. Then, for $v \in W_i$ we obtain

$$0 \leq \lim \Big(\langle v, v\rangle - \frac{1}{f(x_n)^2}\big\langle \exp(M_i x_n) S_i(x_n) v, \exp(M_i x_n) S_i(x_n) v\big\rangle\Big)$$
$$= \langle v, v\rangle - \langle S_i(x_0)v, S_i(x_0)v\rangle,$$

where $\lim \frac{1}{f(x_n)} = 1$ since f is right-continuous at 0. Choose $(\tilde{x}_n)$ such that $\lim \tilde{x}_n = 0$ and $\lim \eta_i(\tilde{x}_n) = -\theta_0$. Then, since $S(x_0)v \in W_i$, we obtain

$$0 \leq \lim \Big(\langle S(x_0)v, S(x_0)v\rangle - \frac{1}{f(\tilde{x}_n)^2}\big\langle \exp(M_i\tilde{x}_n) S_i(\tilde{x}_n) S_i(x_0) v, \exp(M_i\tilde{x}_n) S_i(\tilde{x}_n) S_i(x_0) v\big\rangle\Big)$$
$$= \langle S_i(x_0)v, S_i(x_0)v\rangle - \langle v, v\rangle,$$

since $Q_{d_i}(-\theta_0) Q_{d_i}(\theta_0) = \mathrm{id}$. Hence we obtain

$$\langle v, v\rangle = \langle S(x_0)v, S(x_0)v\rangle,$$

which implies that $S(x_0)$ is an isometry on W_i. Next we show that all W_i are pairwise orthogonal. Let $v \in W_i$ and $u \in W_j$ for $i \neq j$. By Lemma 3.A.3, w.l.o.g. there is a sequence $\hat{x}_n$ such that $\lim \hat{x}_n = 0$, $\lim \eta_i(\hat{x}_n) = \pi$ and $\lim \eta_j(\hat{x}_n) = \theta \in (-\pi, \pi)$. Hence $\lim S(\hat{x}_n)|_{W_i} = -\mathrm{id}$ and $\lim S(\hat{x}_n)|_{W_j} = A_j Q_{d_j}(\theta) A_j^{-1}$ with $\theta \in (-\pi, \pi)$. We have the identity $Q_{d_j}(\phi) + Q_{d_j}(-\phi) = 2\cos(\phi)\,\mathrm{id}$. Applying this to $\phi := \theta/2$ we obtain

$$\mathrm{id} + Q_{d_j}(\theta) = 2\cos(\theta/2) Q_{d_j}(\theta/2).$$

Hence we have $(\mathrm{id} + Q_{d_j}(\theta))^{-1} = (2\cos(\theta/2))^{-1} Q_{d_j}(-\theta/2)$. Set $\tilde{u} := A_j^{-1}(\mathrm{id} + Q_{d_j}(\theta))^{-1} A_j u$. Then $\tilde{u} \in W_j$, and applying (3.5.3) to $v + \tilde{u}$ and taking the limit along $\hat{x}_n$ yields

$$\begin{aligned}
0 &\leq \langle v+\tilde{u}, v+\tilde{u}\rangle - \langle \lim S(\hat{x}_n)(v+\tilde{u}), \lim S(\hat{x}_n)(v+\tilde{u})\rangle \\
&= \|v\|^2 + \|\tilde{u}\|^2 - \|\lim S(\hat{x}_n)v\|^2 - \|\lim S(\hat{x}_n)\tilde{u}\|^2 + 2\langle v, \tilde{u}\rangle - 2\langle \lim S(\hat{x}_n)v, \lim S(\hat{x}_n)\tilde{u}\rangle \\
&= 2\langle v, \tilde{u}\rangle - 2\langle \lim S(\hat{x}_n)v, \lim S(\hat{x}_n)\tilde{u}\rangle,
\end{aligned}$$

where the last equality follows from the fact that each $S(\hat{x}_n)$ is an isometry on W_i and W_j. Since the inequality above also holds for $v - \tilde{u}$, we obtain the equality

$$\begin{aligned}
0 =& \langle v, \tilde{u}\rangle - \langle \lim S(\hat{x}_n)v, \lim S(\hat{x}_n)\tilde{u}\rangle \\
=& \langle v, \tilde{u}\rangle + \langle v, A_j Q_{d_j}(\theta) A_j^{-1}\tilde{u}\rangle = \langle v, A_j(\mathrm{id} + Q_{d_j}(\theta))A_j^{-1}\tilde{u}\rangle = \langle v, u\rangle.
\end{aligned}$$

Since v, u were arbitrary, this shows orthogonality. Hence the decomposition $V = \bigoplus_{i=1}^k W_i$ is orthogonal, and since $S_i(x)$ is an isometry on W_i, it follows that $S(x) := \bigoplus_{i=1}^k S_i(x)$ is in $SO(d)$. Setting $M = \bigoplus_{i=1}^k M_i$, we obtain

$$g(x) = S(x)\exp(Mx).$$

□

Corollary 3.5.4. *Assume that $S(x) = AQ_d^\nu(x)A^{-1}$ is a semigroup with A being an invertible matrix such that $S(x)$ is an isometry for each $x \geq 0$. Then there exists an orthogonal matrix U such that*

$$S(x) = UQ_d^\nu(x)U^{\mathrm{T}}.$$

Proof. One can easily verify the identities $S(x) + S(x)^{\mathrm{T}} = 2\cos(\nu(x))\,\mathrm{id}$ and

$$\sin(\nu(y))S(x) - \sin(\nu(x))S(y) = \sin(\nu(x) - \nu(y))\,\mathrm{id}.$$

Choose $x > 0$ such that $\sin(\nu(x)) \neq 0$. Such an x clearly exists since ν is linear on $\mathbb{Q}x$. Choose any $v \in V$ of unit length and set

$$u := \frac{S(x)v - \cos(\nu(x))v}{\sin(\nu(x))}.$$

The definition of u is invariant under the choice of x as long as $\sin(\nu(x)) \neq 0$. This can be seen by noticing that

$$\begin{aligned} &\sin(\nu(y))S(x) - \sin(\nu(x))S(y) = \sin(\nu(x) - \nu(y))\,\mathrm{id} \\ \iff \quad &\frac{S(x)v - \cos(\nu(x))v}{\sin(\nu(x))} = \frac{S(y)v - \cos(\nu(y))v}{\sin(\nu(y))}. \end{aligned}$$

Hence we obtain $S(x)v = \cos(\nu(x))v + \sin(\nu(x))u$. We have

$$\langle u, v\rangle = \frac{1}{\sin(\nu(x))}\langle S(x)v - \cos(\nu(x))v, v\rangle = \frac{1}{\sin(\nu(x))}\big(\langle S(x)v, v\rangle - \cos(\nu(x))\big) = 0,$$

where the last equality follows from

$$2\langle S(x)v, v\rangle = \langle S(x)v, v\rangle + \langle S(x)^{\mathrm{T}}v, v\rangle = \langle S(x)v + S(x)^{\mathrm{T}}v, v\rangle = 2\cos(\nu(x)).$$

Similarly we can show that u is also of unit length. Set $H := \mathrm{span}(u, v)$. Clearly, H is invariant under $(S(x))_x$ and hence so is $H^\perp$ since each $S(x) \in SO(d)$. In this manner we can construct an orthonormal basis, and we denote by U the matrix associated with this change of basis. Then we have

$$Ug(x)U^{\mathrm{T}} = UAS(x)A^{-1}U^{\mathrm{T}}\exp(UMU^{\mathrm{T}}x) = Q_d^\nu(x)\exp(UMU^{\mathrm{T}}x). \qquad \square$$

Lemma 3.5.5. *Suppose that the semigroup g, acting on V, satisfies Assumption 3.2.5. Let $V = V_1 \oplus V_2$ be the decomposition of Corollary 3.3.8 such that $g|_{V_1}$ is invertible. Then, for any $v \in V_1$ we have*

$$\lim_{x\to 0}\frac{\|g(x)v\|}{\|v\|} = 1.$$

Proof. By Theorem 3.5.3, we have $g(x)|_{V_1} = S(x)\exp(Mx)$. Since $S(x) \in SO(V_1)$, we obtain

$$\langle g(x)v, g(x)v\rangle = \langle \exp(Mx)v, \exp(Mx)\rangle, \quad v \in V_1,$$

which is continuous in x. Hence

$$\lim_{x\to 0}\|g(x)v\| = \lim_{x\to 0}\|\exp(Mx)v\| = \|v\|. \qquad \square$$

Corollary 3.5.6. *Suppose that the semigroup g, acting on V, satisfies Assumption 3.2.5. Then the decomposition from Corollary 3.3.8 is orthogonal.*

Proof. Let $V = V_1 \oplus V_2$ with $g(x)|_{V_1}$ being non-degenerate and $g(x)|_{V_2} \equiv 0$. Assume V_1 is not orthogonal to V_2. Then there exists $v \in V_1$ such that $p_{V_2}(v) \neq 0$, where p_{V_2} denotes the orthogonal projection onto V_2. By Lemma 3.5.5, we have

$$\lim_{x\to 0}\frac{\|g(x)v\|}{\|v\|} = 1.$$

Calculating the same limit for $v - p_{V_2}(v)$ we obtain

$$\lim_{x\to 0} \frac{\|g(x)(v - p_{V_2}(v))\|}{\|v - p_{V_2}(v)\|} = \lim_{x\to 0} \frac{\|g(x)v\|}{\|v\|} \frac{\|v\|}{\|v - p_{V_2}(v)\|} = \frac{\|v\|}{\|v - p_{V_2}(v)\|}.$$

Since $\|v\|^2 = \|v - p_{V_2}(v)\|^2 + \|p_{V_2}(v)\|^2$ and $\|p_{V_2}(v)\| \neq 0$, we have $\frac{\|v\|}{\|v - p_{V_2}(v)\|} > 1$. Hence

$$\lim_{x\to 0} \frac{\|g(x)(v - p_{V_2}(v))\|}{\|v - p_{V_2}(v)\|} > 1,$$

but this contradicts $\|g(x)\|_{\mathrm{op}} \leq f(x)$. Hence $V_1 \perp V_2$. □

Proof of Theorem 3.2.6. By Corollary 3.5.6, we have the orthogonal decomposition $V = \tilde{V}_1 \oplus \tilde{V}_2$ with $g(x)|_{\tilde{V}_2} \equiv 0$ and $g(x)|_{\tilde{V}_1}$ non-degenerate. Applying Theorem 3.5.3 to $g(x)|_{\tilde{V}_1}$ yields the result. □

3.A. Cauchy's functional equation

It is classical that all continuous solutions of the equation (3.1.3), $f(x)+f(y) = f(x+y)$, are linear, and that the non-linear solutions are not continuous, even not Lebesgue measurable, and have dense graphs. For this, and further references, we refer to [16, Section 1.1]. In this section, we provide two auxiliary results on Cauchy's equation. They concern lifting solutions from an interval to the real line, resp. the joint behavior of two solutions that differ by a non-linear function.

Lemma 3.A.1. *Let $f : \mathbb{R} \to \mathbb{R}/[-a, a)$ be a solution to Cauchy's functional equation (3.1.3) on $\mathbb{R}/[-a, a)$ with $a > 0$. Then there exists a solution $\tilde{f} : \mathbb{R} \to \mathbb{R}$ of (3.1.3) such that $f(x) \equiv \tilde{f}(x) \mod [-a, a)$. The solution $\tilde{f}$ is linear if and only if f is linear.*

Proof. Take a Hamel basis $(r_i)_{i\in I}$ of $\mathbb{R}$ such that $r_i \in [-a, a)$ for every $i \in I$. This is clearly possible by rescaling every basis element if necessary. For any $x \in \mathbb{R}$ there exists a finite subset $I_x \subset I$ and $c_i \in \mathbb{Q}$ such that $x = \sum_{i\in I_x} c_i r_i$. The function $\tilde{f}(x) = \sum_{i\in I_x} c_i f(r_i)$ satisfies $f(x) \equiv \tilde{f}(x) \mod [-a, a)$. If f is linear, then clearly $\tilde{f}$ is linear as well. If f is not linear then there exist two basis elements r_1 and r_2 such that $f(r_1) - \frac{r_1}{r_2} f(r_2) \neq 0$, and hence $\tilde{f}$ is also not linear. □

Definition 3.A.2. *We say that two solutions ν and η of Cauchy's functional equation are equivalent if $\nu - \eta$ is linear.*

Lemma 3.A.3. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be two non-equivalent solutions of Cauchy's functional equation. Then there exists a sequence $(x_n)_{n\geq 0}$ in $\mathbb{R}_{\geq 0}$, converging to 0, such that either $\lim_n f(x_n) = \pi$ and $\lim_n g(x_n) = \theta$ with $|\theta| < \pi$ or vice versa.*

Proof. Choose $x, y \in \mathbb{R}_{>0}$ such that the two vectors $v_1 = (x, f(x), g(x))$ and $v_2 = (y, f(y), g(y))$ are linearly independent and $\frac{f(x)}{x} - \frac{f(y)}{y} \neq \frac{g(x)}{x} - \frac{g(y)}{y}$. This is possible, since

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

for all $x, y > 0$ would imply that $f - g$ is linear. Assume w.l.o.g. that $\left|\frac{f(x)}{x} - \frac{f(y)}{y}\right| > \left|\frac{g(x)}{x} - \frac{g(y)}{y}\right|$. Since f and g are both linear on $\mathbb{Q}x$ and $\mathbb{Q}y$ and v_1 and v_2 are linearly independent, there exist sequences $q_n \in \mathbb{Q}$ and $r_n \in \mathbb{Q}$ such that $q_n x - r_n y \geq 0$ for every n, $\lim q_n x - r_n y = 0$ and $\lim_n f(q_n x - r_n y) = \lim_n q_n f(x) - r_n f(y) = \pi$. We show that $x_n = q_n x - r_n y$ has the desired property. Clearly $\lim \frac{r_n}{q_n} = \frac{x}{y}$ and

$$\pi = \lim_n q_n f(x) - r_n f(y) = \lim_n q_n \left(f(x) - \frac{r_n}{q_n} f(y) \right) = \left(f(x) - \frac{x}{y} f(y) \right) \lim q_n.$$

Hence

$$\begin{aligned} |\lim q_n g(x) - r_n g(y)| &= \left| \left(g(x) - \frac{x}{y} g(y) \right) \lim q_n \right| \\ &= \left| \pi \left(\frac{g(x)}{x} - \frac{g(y)}{y} \right) / \left(\frac{f(x)}{x} - \frac{f(y)}{y} \right) \right| < \pi. \end{aligned}$$

□

4. The Fan–Taussky–Todd inequalities and the Lumer–Phillips theorem

We argue that a classical inequality due to Fan, Taussky and Todd (1955) is equivalent to the dissipativity of a Jordan block. As the latter can be characterised via the zeros of Chebyshev polynomials, we obtain a short new proof of the inequality. Three other related inequalities are reproven similarly. By the Lumer–Phillips theorem, the semigroup defined by the Jordan block is contractive. This yields new extensions of the classical Fan–Taussky–Todd inequalities. As applications, we give an estimate for the partial sums of a Bessel function, and a contribution to the classification of self-similar Gaussian Markov processes.

4.1. Introduction

In 1955, Fan, Taussky and Todd proved the following two theorems (use the identity $\cos 2\theta = 1 - 2\sin^2\theta$ to get the precise form stated in [36]):

Theorem 4.1.1. *[36, Theorem 9] For real numbers $a_1, \dots, a_n$, with $a_0 := a_{n+1} := 0$, we have*

$$\sum_{k=1}^{n+1} (a_k - a_{k-1})^2 \geq 2\Big(1 - \cos\frac{\pi}{n+1}\Big) \sum_{k=1}^{n} a_k^2. \tag{4.1.1}$$

Theorem 4.1.2. *[36, Theorem 8] For real numbers $a_1, \dots, a_n$, with $a_0 := 0$, we have*

$$\sum_{k=1}^{n} (a_k - a_{k-1})^2 \geq 2\Big(1 - \cos\frac{\pi}{2n+1}\Big) \sum_{k=1}^{n} a_k^2. \tag{4.1.2}$$

The following converse inequalitites are special cases of results by Milovanović and Milovanović [56].

Theorem 4.1.3. *For real numbers $a_1, \dots, a_n$, with $a_0 := a_{n+1} := 0$, we have*

$$\sum_{k=1}^{n+1} (a_k - a_{k-1})^2 \leq 2\Big(1 + \cos\frac{\pi}{n+1}\Big) \sum_{k=1}^{n} a_k^2. \tag{4.1.3}$$

Theorem 4.1.4. *For real numbers $a_1, \dots, a_n$, with $a_0 := 0$, we have*

$$\sum_{k=1}^{n} (a_k - a_{k-1})^2 \leq 2\Big(1 - \cos\frac{2\pi}{2n+1}\Big) \sum_{k=1}^{n} a_k^2. \tag{4.1.4}$$

Alzer [7] gave short proofs of Theorems 4.1.3 and 4.1.4. In this note we give new short proofs of Theorems 4.1.1 and 4.1.3, and analogous proofs of Theorems 4.1.2 and 4.1.4. There are several proofs in the literature [54, 62, 77]. Our approach is symmetric in the sense that our proofs of Theorems 4.1.1 and 4.1.3 are trivial modifications of each other, and the same holds for Theorems 4.1.2 and 4.1.4. The proofs are based on the dissipativity of Jordan blocks. Besides reproving inequalities which are already known (Section 4.2), we note in Section 4.3 that dissipativity characterizes contractiveness of the matrix semigroup generated by the Jordan block, by the Lumer–Phillips theorem. This leads to new generalizations of Theorems 4.1.1 and 4.1.2. These generalizations are then applied to estimate partial sums of the modified Bessel function of the first kind of order zero. In Section 4.4, we prove strict versions of the new inequalities. Section 4.5 presents an application of dissipativity of Jordan blocks to the classification of self-similar Gaussian Markov processes.

4.2. Proofs

Define the Jordan block

$$J_n(x) := \begin{pmatrix} x & 1 & & 0 \\ & x & 1 & \\ & & \ddots & 1 \\ 0 & & & x \end{pmatrix} \in \mathbb{R}^{n\times n}, \quad x \in \mathbb{R}.$$

We denote the standard scalar product on $\mathbb{R}^n$ by $\langle\cdot,\cdot\rangle$, and use the following terminology from operator theory:

Definition 4.2.1. *(see [27, p. 52]) A matrix $A \in \mathbb{R}^{n\times n}$, not necessarily symmetric, is* dissipative, *if $\langle Aa, a\rangle \le 0$ for all $a \in \mathbb{R}^n$.*

Since

$$\langle J_n(\alpha)a, a\rangle = \alpha \sum_{k=1}^{n} a_k^2 + \sum_{k=2}^{n} a_k a_{k-1}, \quad a = (a_1,\dots,a_n)^{\mathrm{T}} \in \mathbb{R}^n,$$

it is clear that (4.1.1) and (4.1.3) are immediate consequences of the following lemma.

Lemma 4.2.2. *Let $\alpha \in \mathbb{R}$.*
(i) $J_n(\alpha)$ is dissipative if and only if $\alpha \le -\cos(\pi/(n+1))$.
(ii) $-J_n(\alpha)$ is dissipative if and only if $\alpha \ge \cos(\pi/(n+1))$.

Proof. Since $2a^{\mathrm{T}}J_n(\alpha)a = a^{\mathrm{T}}J_n(\alpha)^{\mathrm{T}}a + a^{\mathrm{T}}J_n(\alpha)a$, $a \in \mathbb{R}^n$, the first condition in (i) is equivalent to $B_n(\alpha) := J_n(\alpha)^{\mathrm{T}} + J_n(\alpha)$ being negative semidefinite, i.e. all its eigenvalues being non-positive. It is well-known [23, pp. 25-26] that $\det B_n(x) = U_n(x)$, $x \in \mathbb{R}$, where $U_n(x)$ is the nth Chebyshev polynomial of the second kind. Thus, any eigenvalue μ

of $B_n(\alpha)$ must satisfy

$$0 = \det(B_n(\alpha) - \mu I) = \det B_n(\alpha - \tfrac{1}{2}\mu) = U_n(\alpha - \tfrac{1}{2}\mu).$$

Now note that all such μ are ≤ 0 if and only if

$$\alpha \leq \min_{1\leq k\leq n} \cos\frac{k\pi}{n+1} = \cos\frac{n\pi}{n+1} = -\cos\frac{\pi}{n+1},$$

where $\cos(k\pi/(n+1))$ are the zeros of U_n. The proof of (ii) is analogous; now α must satisfy

$$\alpha \geq \max_{1\leq k\leq n} \cos\frac{k\pi}{n+1} = \cos\frac{\pi}{n+1}. \qquad \square$$

To prove Theorems 4.1.2 and 4.1.4, define

$$\tilde{J}_n(x) := \begin{pmatrix} x & 1 & & & 0 \\ & x & 1 & & \\ & & \ddots & 1 & \\ & & & x & 1 \\ 0 & & & & x - \frac{1}{2} \end{pmatrix} \in \mathbb{R}^{n\times n}, \quad x \in \mathbb{R}.$$

Lemma 4.2.3. *For $x \in \mathbb{R}$ and $n \in \mathbb{N}$, we have the determinant evaluation*

$$\det(\tilde{J}_n(x)^{\mathrm{T}} + \tilde{J}_n(x)) = U_n(x) - U_{n-1}(x).$$

As above, $U_n(x)$ denotes the nth Chebyshev polynomial of the second kind.

Proof. By a classical result on tridiagonal matrices [57, Chapter XIII], this determinant is $b_n = b_n(x)$, where the sequence $(b_k)_{-1\leq k\leq n}$ is defined by $b_{-1} = 0$, $b_0 = 1$, and

$$\begin{aligned} b_k &= 2xb_{k-1} - b_{k-2}, \quad 1 \leq k < n, \\ b_n &= (2x-1)b_{n-1} - b_{n-2}. \end{aligned}$$

For $k < n$, this is the recurrence for the Chebyshev polynomials of the second kind, and so $b_k = U_k$ for $k < n$. Finally,

$$b_n = (2x-1)U_{n-1} - U_{n-2} = U_n - U_{n-1}. \qquad \square$$

Lemma 4.2.4. *For $n \in \mathbb{N}$, the zeros of $U_n(x) - U_{n-1}(x)$ are $(-1)^{k+1}\cos(k\pi/(2n+1))$, $1 \leq k \leq n$.*

Proof. Let $k \in \{1, \ldots, n\}$ be even. We have

$$-\cos\frac{k\pi}{2n+1} = \cos\Big(\pi - \frac{k\pi}{2n+1}\Big).$$

The assertion now follows from the representation

$$U_n(\cos\theta) = \frac{\sin((n+1)\theta)}{\sin\theta}, \quad \theta \in \mathbb{R},$$

because

$$\begin{aligned} n\Big(\pi - \frac{k\pi}{2n+1}\Big) &= \Big(n - \frac{k}{2} + \frac{1}{2}\Big)\pi - \frac{(2n-k+1)\pi}{4n+2}, \\ (n+1)\Big(\pi - \frac{k\pi}{2n+1}\Big) &= \Big(n - \frac{k}{2} + \frac{1}{2}\Big)\pi + \frac{(2n-k+1)\pi}{4n+2}. \end{aligned}$$

The proof for odd k is analogous. □

Since

$$\min_{1\le k\le n} (-1)^{k+1} \cos\frac{k\pi}{2n+1} = -\cos\frac{2\pi}{2n+1}$$

and

$$\max_{1\le k\le n} (-1)^{k+1} \cos\frac{k\pi}{2n+1} = \cos\frac{\pi}{2n+1},$$

Lemmas 4.2.3 and 4.2.4 show that the following result can be proven analogously to Lemma 4.2.2.

Lemma 4.2.5. *Let $\alpha \in \mathbb{R}$.*
(i) $\tilde{J}_n(\alpha)$ is dissipative if and only if $\alpha \le -\cos(2\pi/(2n+1))$.
(ii) $-\tilde{J}_n(\alpha)$ is dissipative if and only if $\alpha \ge \cos(\pi/(2n+1))$.

For $a \in \mathbb{R}^n$, we have

$$a^{\mathrm{T}}\tilde{J}_n(\alpha)a = \alpha\sum_{k=1}^{n} a_k^2 - \frac{a_n^2}{2} + \sum_{k=2}^{n} a_k a_{k-1}.$$

Now (4.1.2) and (4.1.4) easily follow, by applying Lemma 4.2.5 (i), (ii) with the respective optimal values of α.

4.3. A generalization

In what follows, we consider the Euclidean norm on $\mathbb{R}^n$, and write

$$\|A\|_{\mathrm{op}} = \sup_{0\ne a\in\mathbb{R}^n} \frac{\|Aa\|}{\|a\|}$$

for the operator norm of an $n\times n$ matrix A. The contractivity of matrix semigroups w.r.t. this norm is characterized by the Lumer–Phillips theorem [27, p. 52]. In our setting, i.e. for semigroups on $\mathbb{R}^n$, it is the following statement.

Theorem 4.3.1. *For any real $n\times n$ matrix Q, the following are equivalent:*

(i) $\|\exp(Qx)\|_{\text{op}} \leq 1$ *for all* $x \geq 0$,
(ii) Q *is dissipative.*

This yields a generalization of Theorem 4.1.1, involving an extra parameter $x \geq 0$.

Theorem 4.3.2. *For real numbers* $a_1, \ldots, a_n$ *and* $x \geq 0$, *we have*

$$\sum_{j=0}^{n-1} \left(\sum_{k=0}^{j} \frac{x^k}{k!} a_{n-j+k} \right)^2 \leq \exp\left(2x \cos\left(\frac{\pi}{n+1}\right)\right) \sum_{j=1}^{n} a_j^2. \tag{4.3.1}$$

Proof. By Lemma 4.2.2, $J_n(\alpha)$ satisfies condition (ii) of Theorem 4.3.1 for $\alpha = -\cos(\pi/(n+1)$, and so

$$\|\exp(J_n(\alpha)x)a\|^2 \leq \|a\|^2, \quad x \geq 0, \ a \in \mathbb{R}^n. \tag{4.3.2}$$

By calculating the matrix exponential of a nilpotent matrix (see [35, p. 9]),

$$\exp(x(J_n(1) - I)) = \exp \begin{pmatrix} 0 & x & & 0 \\ & 0 & x & \\ & & \ddots & x \\ 0 & & & 0 \end{pmatrix} = \begin{pmatrix} 1 & x & \frac{x^2}{2} & \cdots & \frac{x^{n-1}}{(n-1)!} \\ & 1 & x & \cdots & \frac{x^{n-2}}{(n-2)!} \\ & & \ddots & & \\ & & & & 1 \end{pmatrix}, \quad x \in \mathbb{R},$$

we find

$$\exp(J_n(\alpha)x) = \exp(x(J_n(1) - I) + \alpha x I) = e^{\alpha x} \begin{pmatrix} 1 & x & \frac{x^2}{2} & \cdots & \frac{x^{n-1}}{(n-1)!} \\ & 1 & x & \cdots & \frac{x^{n-2}}{(n-2)!} \\ & & \ddots & & \\ & & & & 1 \end{pmatrix}. \tag{4.3.3}$$

Using this in (4.3.2) yields (4.3.1), because

$$\begin{aligned} \|\exp(J_n(\alpha)x)a\|^2 &= e^{2\alpha x} \sum_{m=1}^{n} \left(\sum_{k=m}^{n} a_k \frac{x^{k-m}}{(k-m)!} \right)^2 \\ &= e^{2\alpha x} \sum_{j=0}^{n-1} \left(\sum_{k=n-j}^{n} a_k \frac{x^{k+j-n}}{(k+j-n)!} \right)^2 = e^{2\alpha x} \sum_{j=0}^{n-1} \left(\sum_{k=0}^{j} a_{n-j+k} \frac{x^k}{k!} \right)^2. \square \end{aligned}$$

As (4.3.1) is obviously sharp for $x \downarrow 0$, the inequality must hold after taking the derivative w.r.t. x at zero on both sides. An easy calculation shows that this yields (4.1.1), and so Theorem 4.3.2 can be viewed as a generalization of Theorem 4.1.1. Another special case might be worth mentioning: Putting $a = (0, \ldots, 0, 1)$ in (4.3.1) yields

$$\sum_{j=0}^{n-1} \frac{x^{2j}}{j!^2} \leq \exp\left(2x \cos\left(\frac{\pi}{n+1}\right)\right), \quad n \in \mathbb{N}, \ x \geq 0. \tag{4.3.4}$$

This inequality for the partial sums of the modified Bessel function of the first kind $I_0(2x) = \sum_{j=0}^{\infty} \frac{x^{2j}}{j!^2}$ seems to be new. Analogously to Theorem 4.3.2, we can generalize Theorem 4.1.2 as follows.

Theorem 4.3.3. *For real numbers $a_1, \ldots, a_n$ and $x \geq 0$, we have*

$$\sum_{j=1}^{n-1} \left(\sum_{k=0}^{j} \frac{x^k}{k!} a_{n-j+k} \right)^2 + e^{-x} a_n^2 \leq \exp\left(2x \cos\left(\frac{2\pi}{2n+1}\right)\right) \sum_{j=1}^{n} a_j^2.$$

Proof. The proof is very similar to that of Theorem 4.3.2. Use Lemma 4.2.5, with $\alpha = -\cos(2\pi/(2n+1)$, and calculate $\|\exp(\tilde{J}_n(\alpha)x)a\|^2$ in order to apply Theorem 4.3.1 . ☐

Taking the derivative of (4.3.1) at zero yields (4.1.2) (cf. the remark after Theorem 4.3.2). For $a = (0, \ldots, 0, 1)$, as above, (4.3.1) yields

$$\sum_{j=0}^{n-1} \frac{x^{2j}}{j!^2} \leq 1 - e^{-x} + \exp\left(2x \cos\left(\frac{2\pi}{2n+1}\right)\right), \quad n \in \mathbb{N},\ x \geq 0. \tag{4.3.5}$$

For fixed n, the estimate (4.3.5) is sharper than (4.3.4) for large x. We conjecture that there is a threshold $x_0 = x_0(n) > 0$ such that (4.3.4) gives a better bound for $0 < x < x_0$.

4.4. Strict inequalities

We were not able the find the following strict version of the Lumer–Phillips theorem in the literature:

Theorem 4.4.1. *For any real $n \times n$ matrix Q, the following are equivalent:*
(i) $\|\exp(Qx)\|_{\mathrm{op}} < 1$ *for all* $x > 0$,
(ii) $\langle Qa, a \rangle < 0$ *for all* $a \in \mathbb{R}^n \setminus \{0\}$.

Analogously to Theorem 4.3.2, Theorem 4.4.1 implies the following statement, and an analogous strict variant of Theorem 4.3.3.

Theorem 4.4.2. *For real numbers $a_1, \ldots, a_n$, not all zero, $x > 0$ and $\alpha > \cos(\pi/(n+1))$, we have*

$$\sum_{j=0}^{n-1} \left(\sum_{k=0}^{j} \frac{x^k}{k!} a_{n-j+k} \right)^2 < e^{2x\alpha} \sum_{j=1}^{n} a_j^2.$$

It remains to prove Theorem 4.4.1.

Lemma 4.4.3. *Let Q be a real $n \times n$ matrix satisfying $\|\exp(Qx)\|_{\mathrm{op}} \leq 1$ for all $x \geq 0$. Assume there exists $x_0 > 0$ such that $\|\exp(Qx_0)\|_{\mathrm{op}} = 1$. Then $\|\exp(Qx)\|_{\mathrm{op}} = 1$ for all $x \geq 0$.*

Proof. Choose $a \in \mathbb{R}^n$ with unit length such that $\|\exp(Qx_0)a\| = 1$. Clearly we must have $\|\exp(Qx)a\| = 1$ for all $x \in [0, x_0]$, since otherwise, by submultiplicativity,

$$\begin{aligned} \|\exp(Qx_0)a\| &= \|\exp(Q(x_0 - x))\exp(Qx)a\| \\ &\leq \underbrace{\|\exp(Q(x_0 - x))\|}_{\leq 1} \underbrace{\|\exp(Qx)a\|}_{<1} < 1. \end{aligned}$$

Now $\langle \exp(Qx)a, \exp(Qx)a\rangle - 1$ is analytic in x and vanishes on $[0, x_0]$. Hence it must vanish everywhere, and we have $\|\exp(Qx)a\| = 1$ for all x. □

Lemma 4.4.4. *Let Q be a real $n \times n$ matrix satisfying $\|\exp(Qx)\|_{\mathrm{op}} \leq 1$ for all $x \geq 0$. The set of vectors $a \in \mathbb{R}^n$ satisfying $\|\exp(Qx)a\| = \|a\|$ for all $x \geq 0$ is a subspace. Denote this subspace by W_Q. Then W_Q and $W_Q^\perp$ are invariant under* $\exp(Qx)$.

Proof. For $a \in \mathbb{R}^n$ and $x \geq 0$, we have

$$\|\exp(Qx)a\| = \|a\| \quad \Longleftrightarrow \quad \big\langle (\mathrm{id} - \exp(Q^{\mathrm{T}}x)\exp(Qx))a, a\big\rangle = 0.$$

Our assumption implies that $\mathrm{id} - \exp(Q^{\mathrm{T}}x)\exp(Qx)$ is positive semidefinite, and thus the condition $\|\exp(Qx)a\| = \|a\|$ is equivalent to $a \in \ker(\mathrm{id} - \exp(Q^{\mathrm{T}}x)\exp(Qx))$. By Lemma 4.4.3, this kernel is the same for every $x > 0$. Let $u \in W_Q^\perp$. Then, for any $w \in W_Q$ we have

$$\langle \exp(Qx)u, w\rangle = \langle u, \exp(Qx)^{\mathrm{T}}w\rangle = \langle u, \exp(-Qx)w\rangle = 0,$$

since $\exp(-Qx)w \in W_Q$. Hence $W_Q^\perp$ is invariant under $\exp(Qx)$. Invariance of W_Q is clear. □

Proof of Theorem 4.4.1. If (ii) holds, then $\|\exp(Qx)\|_{\mathrm{op}} \leq 1$, $x \geq 0$, by Theorem 4.3.1. Assume for the sake of contradiction that $\|\exp(Qx)\|_{\mathrm{op}} = 1$ for some $x > 0$. By Lemmas 4.4.3 and 4.4.4, there exists $0 \neq a \in \mathbb{R}^n$ such that $\langle \exp(Qx)a, \exp(Qx)a\rangle$ is constant in x. Hence the derivative at $x = 0$ has to be zero, and hence $\langle Qa, a\rangle = 0$, contradicting our assumption. Conversely, assume there exists $a \neq 0$ with $\langle Qa, a\rangle = 0$. Then we have $a \in \ker(Q + Q^{\mathrm{T}})$, since $Q + Q^{\mathrm{T}}$ is negative semidefinite. Hence $\exp(Q^{\mathrm{T}}x)a = \exp(-Qx)a$, and we have

$$\begin{aligned} \langle \exp(Qx)a, \exp(Qx)a\rangle &= \langle a, \exp(Q^{\mathrm{T}}x)\exp(Qx)a\rangle \\ &= \langle a, \exp(-Qx)\exp(Qx)a\rangle = \langle a, a\rangle. \end{aligned}$$

Therefore, $\|\exp(Qx)\|_{\mathrm{op}} = 1$ for all $x \geq 0$. □

4.5. Application to Gaussian Processes

In chapter 2, all centered H-self-similar Gaussian Markov (SSGM) processes with values in $\mathbb{R}^n$ are characterized. In particular, we have shown that the matrix-valued covariance function

$$R(s,t) = t^{2H} \exp\big(M \log(t/s)\big), \quad H > 0,\ M \in \mathbb{R}^{n\times n},\ 0 \le s \le t,$$

yields such a process, in the sense of Definition 2.2.7 in chapter 2, if and only if the matrix semigroup $(\exp(Mx))_{x\ge 0}$ satisfies the contractivity condition

$$\|\exp(Mx)\|_{\mathrm{op}} \le e^{-Hx}, \quad x \ge 0. \tag{4.5.1}$$

By Theorem 4.3.1, this is equivalent to the dissipativity of $M + H\,\mathrm{id}$. In the special case where the matrix $M = J_n(\lambda)$ consists of a single Jordan block, we can thus apply Lemma 4.2.2 with $\alpha = \lambda + H$ to conclude:

Proposition 4.5.1. *The covariance function*

$$R(s,t) = t^{2H} \exp\big(J_n(\lambda) \log(t/s)\big), \quad H > 0,\ \lambda \in \mathbb{R},\ 0 \le s \le t, \tag{4.5.2}$$

defines an $\mathbb{R}^n$-valued SSGM process if and only if

$$\lambda + H \le -\cos\frac{\pi}{n+1}.$$

Of course, the matrix exponential in (4.5.2) can be evaluated by (4.3.3).

Part II.

Martingales

5. Martingales with restricted support

We consider martingales with given initial and final marginal distribution whose paths are restricted to a subset of the entire path space. Our main result is the formulation of necessary and sufficient condition guaranteeing the existence of such a measure akin to the convex ordering condition of Strassen [71]. We then apply these conditions to obtain no joint arbitrage conditions in discrete time.

5.1. Introduction

A classical result in martingale theory is Strassen's theorem which clarifies the existence of martingale couplings with marginal constraints. Given two probability distributions μ, ν on $\mathbb{R}^d$, there exists a martingale coupling if and only if $\mathbb{E}^{\mu}[f] \le \mathbb{E}^{\nu}[f]$ for all non negative convex functions $f : \mathbb{R}^d \to \mathbb{R}$, i.e., $\mu \preceq \nu$ in *convex order*. In this chapter we study the existence of martingale couplings, such that given some subset $P \subset (\mathbb{R}^d)^n$ for $n \ge 2$, P has probability one under the coupling. Our main contribution is the introduction of a superlinear operator S_P depending on P and acting on all functions $f : \mathbb{R}^d \to \mathbb{R}$ such that there exists a martingale coupling $\mathbb{Q}$ with marginals μ, ν if and only if

$$\mathbb{E}^{\mu}[S_P f] \le \mathbb{E}^{\nu}[f]$$

for a suitable class of test functions f. In the case when P is the entire space, S_P maps any function to its convex envelope and we recover Strassen's original result.

In the context of mathematical finance we apply this result and consider a price process (S_1, S_2) taking values in $\mathbb{R}^d_{>0}$ and some contract with payoff $L(S_1, S_2) \in \mathbb{R}$ at time $t = 2$ and price C at time $t = 1$. We further assume that interest rates are zero. Let the risk free marginal distributions $S_1 \sim \mu_1$, $S_2 \sim \mu_2$ and $C \sim \mu_C$ be fixed. Given the availability of trading strategies involving static positions in vanilla options on S_1, S_2 and C, as well as dynamic trading in the price process S and the contract, the absence of arbitrage (see [42]) implies the existence of a martingale measure $\mathbb{Q}$ on $\mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$ with the given marginals, such that

$$\mathbb{E}^{\mathbb{Q}}[S_2 | \sigma(S_1, C)] = S_1, \quad \mathbb{E}^{\mathbb{Q}}[L(S_1, S_2) | \sigma(S_1, C)] = C.$$

Notice that any such measure $\mathbb{Q}$ must assign full mass to the set $P = \{((s_1, c), (s_2, L(s_1, s_2))) : s_1, s_2 \in \mathbb{R}^d,\ c \in \mathbb{R}\} \subset \mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$. This observation allows us to apply our developed theory to clarify the existence of such couplings and hence the absence of arbitrage via a Strassen-type result. The motivating result for this problem is the absence of joint

arbitrage in a two-step market with the contract being the VIX as discussed by Guyon in [42]. We extend Guyons's result on no joint arbitrage when considering the VIX and develop related results for a larger class of claims.
The chapter is organized as follows, in Section 5.2 we introduce the necessary notation and state the main results. Section 5.3 presents the application of the theory to the mathematical finance setting. In Section 5.4 most of the theory is developed and in Section 5.5 the proofs for the main results can be found. Section 5.6 presents the proofs of Section 5.3 on the applications to mathematical finance.

5.2. Preliminaries and main results

Throughout this chapter X denotes a d-dimensional vector space unless otherwise specified and

$$Z = X^n = \underbrace{X \times \cdots \times X}_{n}$$

the n-fold product of X where $n \in \mathbb{N}$. By $\overline{\mathbb{R}}$ we denote the extended real numbers $\mathbb{R} \cup \{-\infty, \infty\}$.

For the proofs of the theorems stated in this section we refer to Section 5.4.

Definition 5.2.1. *Denote by $C(X)$ the set of real valued continuous functions on X and by $C_b(X)$ the set of bounded continuous real valued functions. For any $\psi \in C_b(X)$ with $\psi > 0$ set*

$$C_\psi(X) := \{f \in C(X) : \; f\psi \in C_b(X)\}$$

and

$$C_l(X) := \{f \in C(X) : \; \exists K \in \mathbb{R} : f \geq K\},$$

the set of continuous function with a lower bound. Set

$$\mathcal{F}_\psi(X) := \{f : X \to \overline{\mathbb{R}} : \; f \text{ upper semi-continuous and } \exists\, g \in C_\psi(X) : \; f \leq g\}.$$

Denote by $M(X)$ the set of all finite signed Borel measures on X and by $M^1(X)$ the subset of probability measures (and similarly $M(Z)$, $M^1(Z)$). Set

$$M_\psi(X) = \{\mu \in M(X) : \; \mathbb{E}^\mu \left[\frac{1}{\psi}\right] < \infty\}.$$

and set $M^1_\psi(X) = M_\psi(X) \cap M^1(X)$ the subset of probability measures.

Definition 5.2.2. *For $1 \leq i \leq n$ denote $p_i : Z \to X$ the projection onto the i-th component. For $P \subset Z$ set $P_i = (p_1, \ldots, p_i)(P) \subset X^i$*

Definition 5.2.3. *Let $f : X \to \overline{\mathbb{R}}$. We say that f is convex if its epigraph*

$$\operatorname{epi}(f) = \{(x, t) \in X \times \mathbb{R} : \; t \geq f(x)\}$$

is a convex set.

Next we define the operator needed to formulate a Strassen-type result for martingales with restricted support.

Definition 5.2.4. *Let $P \subset Z = X^n = X^{n-1} \times X$ and set for $h : X^{n-1} \times X \to \overline{\mathbb{R}}$*

$U_P^y(h) :=$
$\{g : X \to \overline{\mathbb{R}} : g$ *convex and* $g(x) \le h(y,x)$ *for all* $x \in X$ *with* $(y,x) \in P$ *and* $y \in X^{n-1}\}$.

For $h : X^{n-1} \times X \to \overline{\mathbb{R}}$ define $S_P^n h : X^{n-1} \to \overline{\mathbb{R}}$ by

$$(S_P^n h)(y) := \sup_{g \in U_P^y(h)} g(p_{n-1}(y)).$$

For $f : X \to \overline{\mathbb{R}}$ set

$$S_P f : X \to \mathbb{R},\ x \mapsto \left[(S_{P_2}^2 \circ \cdots \circ S_{P_{n-1}}^{n-1} \circ S_P^n)(f \circ p_n)\right](x) = (S_{P_2}^2(S_{P_3}^3(\cdots S_P^n(f \circ p_n)))(x).$$

Remark 5.2.5. *For any function $h : X^{n-1} \times X \to \overline{\mathbb{R}}$ we have $g \in U_P^y(h)$ if and only if $g \in U_Z^y(h_P)$, where*

$$h_P(y,x) = \begin{cases} h(y,x) & (y,x) \in P, \\ \infty & \text{else} \end{cases}$$

with $y \in X^{n-1}$ and $x \in X$. Hence $S_P^n h = S_Z^n h_P$. It is clear from Definition 5.2.4 with P replaced by the entire space Z that $S_Z^n h_P$ is just the convex envelope relative to X (see Definition 5.4.2) of the function

$$x \mapsto h_P(y,x)$$

evaluated at $x = p_{n-1}(y)$.

Example 3. Suppose that $n = 2$ and $Z = X^2$, then for any function $f : X \to \overline{\mathbb{R}}$ we have

$$\begin{aligned} S_P f(x) &= S_{P_2}^2(f \circ p_2)(x) \\ &= \sup\{g(x) :\ g \text{ convex and } g(y) \le (f \circ p_2)(x,y) \text{ for all } y \text{ with } (x,y) \in P\} \end{aligned}$$

which is equivalent to

$$S_P f(x) = \sup\{g(x) :\ g \text{ convex and } g(y) \le f(y) \text{ for all } y \text{ with } (x,y) \in P\}.$$

If $P = Z$, then $S_P f$ is the convex envelope of f and if $P = \Delta := \{(x,x) : x \in X\}$ then $S_P f = f$.

Definition 5.2.6. *Let $P \subset Z$ and denote by $\mathcal{M}(P) \subset M(Z)$ the set of martingale probability measures on Z such that for all $\mathbb{Q} \in \mathcal{M}(P)$ it holds that $\mathbb{Q}(P) = 1$. Set $\mathcal{M}_\psi(P) = M_{\psi_Z}(Z) \cap \mathcal{M}(P)$. For $\mu, \nu \in M_\psi^1(X)$ set*

$${}_\mu\mathcal{M}_\psi(P) = \{\mathbb{Q} \in \mathcal{M}_\psi(P) : (p_1)_*\mathbb{Q} = \mu\}$$

and

$$^{\nu}\mathcal{M}_\psi(P) = \{\mathbb{Q} \in \mathcal{M}_\psi(P) : (p_n)_*\mathbb{Q} = \nu\}$$

where $(\cdot)_*$ *denotes the pushforward. Set*

$$^{\nu}_{\mu}\mathcal{M}_\psi(P) = {}_{\mu}\mathcal{M}_\psi(P) \cap {}^{\nu}\mathcal{M}_\psi(P).$$

We are now ready to state the main result of this chapter.

Theorem 5.2.7. *Let* $P \subset Z$ *be closed and* $\mu, \nu \in M^1(X)$ *have finite first moments. Then for all* $f \in C(X)$, $S_P f$ *is Borel measurable. Moreover there exists* $\mathbb{Q} \in \mathcal{M}(P)$ *with initial marginal distribution* μ *and final marginal distribution* ν *if and only if for all* $f \in C(X)$ *with* $f \geq 0$ *we have*

$$\mathbb{E}^\mu[S_P f] \leq \mathbb{E}^\nu[f].$$

Remark 5.2.8. *Theorem 5.2.7 does not hold in general if* P *is open. Indeed, set* $X = \mathbb{R}$, $n = 2$ *and* $P := \mathbb{R}^2 \setminus \Delta = \{(x, y) : \ x \neq y\}$. *For any* $f \in C(X)$ *with* $f \geq 0$ *we have*

$$S_P f(x) = \sup\{g(x) : \ g \text{ convex and } g(y) \leq f(y) \text{ for all } y \text{ with } y \neq x\}.$$

Since each g *in the above set satisfies* $g < \infty$, *we either have* $g \equiv -\infty$ *or* g *is real valued. In the latter case since* g *is convex, it is continuous and thus* $g(y) \leq f(y)$ *holds for all* y. *Hence* $S_P f = \mathcal{L}f$ *(see Definition 5.A.5). For any probability measure* μ *on* $\mathbb{R}$ *with finite first moment, the identity coupling* $\mathbb{Q}_\mu = (x \mapsto (x, x))_*(\mu)$ *is supported on* Δ *and hence not in* $\mathcal{M}_\psi(P)$. *Let* $\mathbb{Q}$ *be an arbitrary martingale coupling of* (μ, μ) *and denote by* $d\eta_{\mathbb{Q}} d\mu$ *its disintegration. Then we have by Jensen's inequality*

$$\begin{aligned}
\mathbb{E}^\mu[\sqrt{1 + x^2}] &= \int \sqrt{1 + y^2}\mathbb{Q}(dx, dy) \\
&= \int \sqrt{1 + y^2}\eta_{\mathbb{Q}}(x, dy)\mu(dx) \\
&\geq \int \sqrt{1 + x^2}\mu(dx) = \mathbb{E}^\mu[\sqrt{1 + x^2}].
\end{aligned}$$

Since $\sqrt{1 + x^2}$ *is strictly convex and we must have equality, we obtain* $\eta_{\mathbb{Q}}(x, dy) = \delta_x$ *and thus* $\mathbb{Q}$ *must be the identity coupling. But clearly for any* $f \in C(\mathbb{R})$ *with* $f \geq 0$ *we have*

$$\mathbb{E}^\mu[S_P f] = \mathbb{E}^\mu[\mathcal{L}f] \leq \mathbb{E}^\mu[f]$$

since we have $\mathcal{L}f \leq f$. *Hence choosing* $\mu = \nu$ *we obtain that* μ *satisfies the conditions of Theorem 5.2.7 but there is no martingale coupling of* μ *with* μ *supported on* P.

For the rest of the section set $\psi : X \to \mathbb{R}, \ x \mapsto \frac{1}{1+||x||^t}$ with $t \geq 1$. The set $M_\psi(X)$ is the set of measures which integrate $\frac{1}{\psi}$. Hence for $\psi(x) = \frac{1}{1+||x||^t}$, $M_\psi(X)$ is the set of all measures on X with finite t-th moments.

The following theorem analytically characterizes the range of possible final distributions for a martingale with a fixed initial distribution $\mu \in M^1_\psi(X)$ and whose support lies within a closed set P.

Theorem 5.2.9. *Let $\mu \in M^1_\psi(X)$ and $P \subset Z$ be closed. For all $f \in \mathcal{F}_\psi(X)$, $S_P f$ is Borel measurable and if ${}_\mu\mathcal{M}_\psi(P) \neq \emptyset$*

$$\inf_{\mathbb{Q} \in {}_\mu\mathcal{M}_\psi(P)} \mathbb{E}^{\mathbb{Q}}[f \circ p_n] = \mathbb{E}^\mu[S_P f].$$

From a financial perspective, this result provides the lowest arbitrage-free price of a payoff f in a robust sense, given the initial risk-free marginal distribution and the range P of possible evolutions for the price process.

Remark 5.2.10. *In general the duality in Theorem 5.2.9 does not hold if ${}_\mu\mathcal{M}_\psi(P) = \emptyset$. To see this let $X = \mathbb{R}$, $n = 2$ and $P = \{(x, y) : \ |y| \geq e^{x^2}\}$. Assume μ to be the standard normal distribution, i.e.*

$$\mu(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Let $f(x) := |x|$. Observe that for any $(x, y_1) \in P$ and $(x, y_2) \in P$ such that $ay_1 + by_2 = x$ for non negative a, b with $a + b = 1$ we have for any convex g which satisfies $g(y) \leq f(y)$ for $(x, y) \in P$ that by convexity $g(x) \leq af(y_1) + bf(y_2)$. Thus $S_P f(x) \leq af(y_1) + bf(y_2)$. Hence, since (x, e^{x^2}) and $(x, -e^{x^2})$ are both in P and by the above observation,

$$S_P f(x) \leq af(-e^{x^2}) + bf(e^{x^2}) = e^{x^2},$$

where a and b are nonnegative and satisfy $a(-e^{x^2}) + be^{x^2} = (b - a)e^{x^2} = x$ and $a + b = 1$. Furthermore

$$g(y) = \begin{cases} e^{x^2} & |y| \leq e^{x^2} \\ |y| & \text{else} \end{cases}$$

is convex and satisfies $g(y) \leq f(y)$ for $(x, y) \in P$. Since $g(x) = e^{x^2}$ we obtain $S_P f(x) = e^{x^2}$. It follows that $\mathbb{E}^\mu[S_P f] = \infty$ and hence Theorem 5.2.7 can never be satisfied. Thus there does not exist a martingale measure on P with initial distribution μ. We have $S_P \mathbf{0} = \mathbf{0}$, where $\mathbf{0}(x) \equiv 0$ and hence

$$\mathbb{E}^\mu[S_P \mathbf{0}] = 0 < \infty = \inf_{\mathbb{Q} \in {}_\mu\mathcal{M}_\psi(P)} \mathbb{E}^{\mathbb{Q}}[\mathbf{0} \circ p_2]$$

since ${}_\mu\mathcal{M}_\psi(P)$ is empty.

Definition 5.2.11. *Define for $f : X \to \overline{\mathbb{R}}$*

$$S_P^{-1} f := \{g \in C_l(X) : \ S_P g \geq f\}.$$

Assumption 5.2.12. *For $P \subset Z$ we have for all $f \in C_\psi(X)$ that $S_P f : X \to \overline{\mathbb{R}}$ is lower semi-continuous.*

A sufficient condition for Assumption 5.2.12 is given by the following proposition.

Proposition 5.2.13. *Assumption 5.2.12 holds if P is compact or $P = Z$.*

Under Assumption 5.2.12 we can in a similar fashion as in Theorem 5.2.9 characterize the range of initial marginal distributions of a martingale with given path space P and final distribution $\nu \in M^1_\psi(X)$.

Theorem 5.2.14. *Let $\nu \in M^1_\psi(X)$ and $P \subset Z$ be closed and satisfy Assumption 5.2.12. Assume ${}^\nu\mathcal{M}_\psi(P) \neq \emptyset$. Then for all $f \in \mathcal{F}_\psi(X)$ we have*

$$\sup_{\mathbb{Q} \in {}^\nu\mathcal{M}_\psi(P)} \mathbb{E}^{\mathbb{Q}}[f \circ p_1] = \inf_{g \in S_P^{-1} f} \mathbb{E}^\nu[g].$$

Remark 5.2.15. *This theorem is an extension of Theorem 1 in [29] to all sets satisfying Assumption 5.2.12. In particular for $X = \mathbb{R}^d$, $P = \mathbb{R}^d \times \mathbb{R}^d$, $\nu \in M^1_\psi(\mathbb{R}^d)$ and $g \in C_l(\mathbb{R}^d)$ we have $S_P g = \mathcal{L}g$, where $\mathcal{L}g$ denotes the lower semi-continuous convex envelope of g (see Definition 5.A.5). We obtain*

$$\sup_{\nu \succeq \mu} \mathbb{E}^\mu[f] = \inf_{\substack{\mathcal{L}g \geq f \\ g \in C_l(\mathbb{R}^d)}} \mathbb{E}^\nu[g]$$

for all $f \in \mathcal{F}_\psi(\mathbb{R}^d)$. For any $g \in C_l(\mathbb{R}^d)$ satisfying $\mathcal{L}g \geq f$ it follows that $f \leq \mathcal{L}g \leq g$ and hence $\mathcal{L}g$ is real valued. Together with $\mathcal{L}g$ being convex we obtain that $\mathcal{L}g$ is continuous and in particular by Baire's Theorem for lower semi-continuous functions applied to $g - \mathcal{L}g \geq 0$ (see [72] pg. 79) this implies that there exists a pointwise decreasing sequence $(g_n)_n$ with $g_n \in C_l(\mathbb{R}^d)$ with

$$\lim_{n \to \infty} g_n(x) = \mathcal{L}g(x)$$

pointwise for all $x \in \mathbb{R}^d$. Since $g_n \geq \mathcal{L}g$ we have $\mathcal{L}g_n \geq \mathcal{L}g \geq f$. Thus by monotone convergence

$$\lim_{n \to \infty} \mathbb{E}^\nu[g_n] = \mathbb{E}^\nu[\mathcal{L}g].$$

Since we can construct such a sequence for all $g \in C_l(\mathbb{R}^d)$ this implies

$$\inf_{\substack{\mathcal{L}g \geq f \\ g \in C_l(\mathbb{R}^d)}} \mathbb{E}^\nu[g] \leq \inf_{\substack{\mathcal{L}g \geq f \\ g \in C_l(\mathbb{R}^d)}} \mathbb{E}^\nu[\mathcal{L}g]$$

and the reverse inequality is obvious since $\mathcal{L}g \leq g$. Since for any convex function $g \in C(X)$ we have $\mathcal{L}g = g$ we obtain

$$\sup_{\nu \succeq \mu} \mathbb{E}^\mu[f] = \inf_{\substack{g \geq f \\ g \in C(\mathbb{R}^d) \text{ and convex}}} \mathbb{E}^\nu[g].$$

For functions $f_i : \mathbb{R} \to \mathbb{R}$ with $1 \leq i \leq d$ denote by

$$\bigoplus_{i=1}^d f_i : (X \ni (x_1, \dots, x_n)) \mapsto \sum_{i=1}^d f_i(x_i).$$

Let $\gamma : \mathbb{R} \to \mathbb{R}$ be given by $\gamma(x) = \frac{1}{1+|x|^t}$. In the same way that we provided an existence result for martingale measures with fixed initial and final distributions and a specified path space, we can, under Assumption 5.4.18, establish an analogous result for the existence of a martingale measure with a fixed path space, where the one-dimensional marginal distributions of the initial and final marginal distributions are fixed.

Theorem 5.2.16. *Let $P \subset Z$ be closed and satisfy Assumption 5.2.12 and let $\mu_i, \nu_i \in M^1_\gamma(\mathbb{R})$, $1 \le i \le d$ be probability measures on $\mathbb{R}$. Then there exists a martingale measure $\mathbb{Q} \in \mathcal{M}_\psi(P)$ with initial and final marginal μ and ν whose respective marginals are μ_i and ν_i if and only if for all $f_i \in C(\mathbb{R})$ with $f_i \ge 0$ and $\phi_i \in C_b(\mathbb{R})$ with $\bigoplus_i^d \phi_i \le S_P(\bigoplus_i^d f_i)$ we have*

$$\sum_i \mathbb{E}^{\mu_i}[\phi_i] \le \sum_i \mathbb{E}^{\nu_i}[f_i].$$

In the case that such a $\mathbb{Q}$ exists we have the following duality for all $f_i \in C_\gamma(\mathbb{R})$

$$\sup_{\substack{\bigoplus \phi_i \le S_P(\bigoplus f_i) \\ \phi_i \in L^1(\mu_i)}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] = \inf_{\nu \in \mathcal{M}^n_\psi(\mu_1,\ldots,\mu_d,P)} \sum_i \mathbb{E}^{\nu_i}[f_i].$$

where $\mathcal{M}^n_\psi(\mu_1, \ldots, \mu_d, P) \subset \mathcal{M}_\psi(P)$ is the set of all martingale measures $\mathbb{Q}$ supported on P such that their initial marginal μ has marginals $\mu_1, \ldots, \mu_d$.

5.3. Applications to mathematical finance and no joint arbitrage

Assume we have a $\mathbb{R}^d_{>0}$ valued price process S. Denote by S_1 and S_2 the price at some fixed times $0 < t_1 < t_2$ respectively. We also assume that there is a contract with payoff $L(S_1, S_2)$ at t_2 and price C at t_1 where $L : \mathbb{R}^d_{>0} \times \mathbb{R}^d_{>0} \to \mathbb{R}$. Assume we are given the risk neutral distributions of $S_1 \sim \mu_1$, $C \sim \mu_C$ and $S_2 \sim \mu_2$ from market data and assume we have no interest rate.

Definition 5.3.1. *We call a semi-static trading strategy in (S_1, S_2, C) any strategy of the form*

$$u_1(S_1) + u_2(S_2) + u_C(C) + h_1(S_2 - S_1) + h_2(L(S_1, S_2) - C)$$

where $u_1 \in L^1(\mu_1), u_2 \in L^1(\mu_2), u_C \in L^1(\mu_C)$ and with the deltas h_1, h_2 being bounded and $\sigma(S_1, C)$-measurable.

For the next definition, see Definition 5 in [42].

Definition 5.3.2 (joint arbitrage)**.** *We say that there exists (S_1, S_2, C)-joint arbitrage if there is a semi-static trading strategy $(u_1, u_2, u_C, h_1, h_2)$ with payoff*

$$u_1(S_1) + u_2(S_2) + u_C(C) + h_1(S_2 - S_1) + h_2(L(S_1, S_2) - C) \ge 0$$

and price

$$\mathbb{E}^{\mu_1}[u_1] + \mathbb{E}^{\mu_2}[u_2] + \mathbb{E}^{\mu_C}[u_C] < 0.$$

Note that this definition differs from the standard 'existence of an equivalent martingale measure' concept of arbitrage, which typically requires a reference or market measure. This alternative notion of no-arbitrage is commonly explored in the context of robust finance; see, for example, [14]. In order to exclude joint arbitrage, by [42] and [14], there must exist a martingale measure $\mathbb{Q}$ such that the process

$$(S_1, C) \to (S_2, L(S_1, S_2))$$

is a martingale.

Definition 5.3.3. *Let $L : \mathbb{R}^d_{>0} \times \mathbb{R}^d_{>0} \to \mathbb{R}$ be continuous. Let $\mathcal{P}_L(\mu_1, \mu_C, \mu_2)$ denote the set of probability measures μ on $\mathbb{R}^d_{>0} \times \mathbb{R} \times \mathbb{R}^d_{>0}$ such that*

$$S_1 \sim \mu_1, \quad C \sim \mu_C, \quad S_2 \sim \mu_2, \quad \mathbb{E}^\mu[S_2|S_1, C] = S_1, \quad \mathbb{E}^\mu[L(S_1, S_2)|S_1, C] = C.$$

We set $P := \{((s_1, c), (s_2, L(s_1, s_2))) : \ s_1, s_2 \in \mathbb{R}^d_{>0}, \ c \in \mathbb{R}\} \subset \mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$ and for $f \in C(\mathbb{R}^d)$ by a slight abuse of notation set $S_P f := S_P \widehat{f}$ where $\widehat{f} \in C(\mathbb{R}^{d+1})$ satisfies

$$\widehat{f}(s_1, c) = f(s_1).$$

Remark 5.3.4. *We equip the space $\mathbb{R}^d_{>0} \times \mathbb{R} \times \mathbb{R}^d_{>0}$ with the subspace topology induced by $\mathbb{R}^{2d+1}$. Notice that $\mathbb{R}^d_{>0}$ with its induced topology is Polish.*

The fact that these spaces are Polish, allows us to apply Strassen's original theorem, theorem 7 in [71]. To work with arbitrary marginal distributions, we require some boundedness on the payoff of the contract to ensure the existence of expectations.

Assumption 5.3.5. *Assume for $L : \mathbb{R}^d_{>0} \times \mathbb{R}^d_{>0} \to \mathbb{R}$ there exists $\psi : \mathbb{R}^d_{>0} \to \mathbb{R}$ such that for all $x, y \in \mathbb{R}^d_{>0}$ we have*

$$|L(x, y)| \leq R\left(\frac{1}{\psi(x)} + \frac{1}{\psi(y)}\right).$$

for some constant $R > 0$.

Note that we here allow for general ψ not necessarily $\frac{1}{1+\|x\|^t}$ as in the previous section.

Assumption 5.3.6. *Fix ψ such that Assumption 5.3.5 is satisfied. Then, we assume that $P = \{((s_1, c), (s_2, L(s_1, s_2))) : \ s_1, s_2 \in \mathbb{R}^d_{>0}, \ c \in \mathbb{R}\} \subset \mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$ has the property that $S_P f$ is lower semi-continuous for all $f \in C_\psi(\mathbb{R}^d_{>0})$.*

The next Theorem provides equivalent conditions for the existence of no arbitrage in the sense of Definition 5.3.2.

Theorem 5.3.7. *Assume L satisfies Assumption 5.3.5 with some fixed ψ. Set $P = \{((s_1, c), (s_2, L(s_1, s_2))) : \ s_1, s_2 \in \mathbb{R}^d_{>0}, \ c \in \mathbb{R}\} \subset \mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$ and assume that P is closed and satisfies Assumption 5.3.6. Let $\mu_1, \mu_2 \in M_\psi(\mathbb{R}^d_{>0})$ and assume that μ_1, μ_2, μ_C*

have finite first moments. Then $\mathcal{P}_L(\mu_1, \mu_C, \mu_2) \neq \emptyset$ if and only if for all $f \in C_\psi(\mathbb{R}^d_{>0})$ with $f \geq 0$, $\varphi_1 \in C_b(\mathbb{R}^d_{>0})$ and $\varphi_C \in C_b(\mathbb{R})$ with $S_P f \geq \varphi_1 \oplus \varphi_C$ we have

$$\mathbb{E}^{\mu_1}[\varphi_1] + \mathbb{E}^{\mu_C}[\varphi_C] \leq \mathbb{E}^{\mu_2}[f]. \tag{5.3.1}$$

This allows us to extend Theorem 6 in [42]. Let

$$L(s_1, s_2) = -\frac{2}{\tau}\log(s_2/s_1).$$

Then the price of the contract C is given by VIX^2 at time t_1. Set $P = \{((s_1, c), (s_2, L(s_1, s_2))) : s_1, s_2 \in \mathbb{R}_{>0},\ c \in \mathbb{R}\} \subset \mathbb{R}^2 \times \mathbb{R}^2$ and set

$$\frac{1}{\psi(x)} := 1 + |x| + |\log(x)|.$$

Definition 5.3.8. *For $f \in C_\psi(\mathbb{R})$ set*

$$\widetilde{f}(c) = \sup\{a + bc + d :\ a, b, d \in \mathbb{R},\ ax - b\frac{2}{\tau}\log(x) + d \leq f(x) \text{ for } x > 0\}.$$

Lemma 5.3.9. *For $f \in C_\psi(\mathbb{R})$ we have $\widetilde{f}(0) = f(1)$ and for $c < 0$ we have $\widetilde{f}(0) = \infty$.*

The function $\widetilde{f}$ can be interpreted as follows: If $S_1 = 1$ and c is the price of the contract with payoff $-\frac{2}{\tau}\log(S_2)$ at time $t = 1$ then $\widetilde{f}(c)$ is the maximal subhedging price of a derivative with payoff f at time $t = 2$. Any subhedging strategy where we buy the asset a-times with $S_1 = 1$, the claim b-times at price $c \in \mathbb{R}$ and holding d in the bank account, must satisfy $aS_2 - b\frac{2}{\tau}\log(S_2) + d \leq f(S_2)$ and has cost $a + bc + d$. Taking the supremum over all such strategies yields $\widetilde{f}(c)$.
The primary reason for considering the mapping $f \mapsto \widetilde{f}$ lies in its relationship with the operator S_P. By examining this relationship, we can show that P and L must satisfy Assumption 5.3.5. This yields the following Theorem:

Theorem 5.3.10. *Let μ_1, μ_2 and μ_{VIX^2} have a finite first moment and let $\mathbb{E}^{\mu_i}[|\log(x)|] < \infty$. Then L and P satisfy Assumptions 5.3.5 and 5.3.6 and we have for $f \in C_\psi(\mathbb{R})$.*

$$S_P f(s, c) = \widetilde{f}(c - \frac{2}{\tau}\log(s)).$$

The following are equivalent:

1. *The market is free of joint $(S_1, S_2, \mathrm{VIX}^2)$ arbitrage.*

2. *$\mathcal{P}_L(\mu_1, \mu_{\mathrm{VIX}^2}, \mu_2) \neq \emptyset$.*

3. *For all $f \in C_\psi(\mathbb{R})$ with $f \geq 0$ and $\varphi_1 \in C_b(\mathbb{R}^d_{>0})$ and $\varphi_{\mathrm{VIX}^2} \in C_b(\mathbb{R})$ with $S_P f \geq \varphi_1 \oplus \varphi_{\mathrm{VIX}^2}$ we have*

$$\mathbb{E}^{\mu_1}[\varphi_1] + \mathbb{E}^{\mu_{\mathrm{VIX}^2}}[\varphi_{\mathrm{VIX}^2}] \leq \mathbb{E}^{\mu_2}[f].$$

Fix $\mu_1, \mu_{\mathrm{VIX}^2}$. If we have $\mathcal{P}_L(\mu_1, \mu_{\mathrm{VIX}^2}, \mu_2) \neq \emptyset$ for some μ_2 then we have the following duality for all $f \in C_\psi(\mathbb{R}^d)$

$$\sup_{\substack{\varphi_1 \oplus \varphi_{\mathrm{VIX}^2} \leq S_P f \\ \varphi_1 \in L^1(\mu_1) \\ \varphi_{\mathrm{VIX}^2} \in L^1(\mu_{\mathrm{VIX}^2})}} \mathbb{E}^{\mu_1}[\varphi_1] + \mathbb{E}^{\mu_{\mathrm{VIX}^2}}[\varphi_{\mathrm{VIX}^2}] = \inf_{\substack{\mu_2 \\ \mathcal{P}_L(\mu_1, \mu_{\mathrm{VIX}^2}, \mu_2) \neq \emptyset}} \mathbb{E}^{\mu_2}[f]$$

Consider all contracts on S and VIX with payoffs $\varphi_1(S_1)$ and $\varphi_{\mathrm{VIX}^2}(\mathrm{VIX}^2)$ at time $t = t_1$ such that their joint payoff is less than the maximal subreplication price $S_P f(S_1, \mathrm{VIX}^2)$ of $f(S_2)$. The duality result above states that the supremum of the joint price over all such contracts is precisely the minimal arbitrage free cost of the contract with payoff $f(S_2)$.

5.4. Theoretical results

5.4.1. Discrete representation of S_P

We derive a dual formulation of the operator S_P. The arguments presented are purely combinatorial and do not rely on topological considerations. Given the focus of this thesis, we frame everything in the terminology of probability theory, although all results can be interpreted as outcomes concerning finite convex combinations in finite-dimensional vector spaces.

Definition 5.4.1. *Let $A \subset X$ and set*

$$\Pi_A := \{\lambda |\ \lambda \textit{ probability measure on } X \textit{ with barycenter and } \lambda(A) = 1\}$$

and denote by $\pi(\lambda)$ the barycenter of λ.

Definition 5.4.2 (Relative lower convex envelope)**.** *For $f : X \to \overline{\mathbb{R}}$ and any $A \subset X$ set for every $x \in X$:*

$$\overline{f}_A(x) := \sup\{g(x) |\ g : X \to \overline{\mathbb{R}} \textit{ convex and } \forall y \in A : g(y) \leq f(y)\}.$$

Proposition 5.4.3. *For any $A \subset X$ and $f : X \to \overline{\mathbb{R}}$ the function $\overline{f}_A(x)$ is convex. We have*

$$\overline{f}_A(x) = \inf_{\lambda \in \Xi^x_A(f)} \mathbb{E}^\lambda[f]$$

where

$$\Xi^x_A(f) = \{\lambda \in \Pi_A :\ \pi(\lambda) = x \quad \textit{and } \operatorname{supp}(\lambda) \textit{ is finite and } f(y) < \infty \textit{ for } y \in \operatorname{supp}(\lambda)\}$$

Proof. For any function f set $\operatorname{epi}(f) := \{(x,t) \in X \times \mathbb{R} : \; t \geq f(x)\}$. Set $M = \{g : X \to \overline{\mathbb{R}} : g$ is convex and $g \leq f$ on A$\}$. Then we have

$$\operatorname{epi}(\overline{f}_A) = \bigcap_{g \in M} \operatorname{epi}(g)$$

and since the intersection of convex sets is convex it follows that $\operatorname{epi}(\overline{f}_A)$ is convex and hence $\overline{f}_A$ is convex. Set

$$H(x) := \inf_{\lambda \in \Xi_A^x(f)} \mathbb{E}^\lambda[f].$$

Choose $(x,t), (y,s) \in \operatorname{epi}(H)$. It follows that

$$t \geq \inf_{\lambda \in \Xi_A^x(f)} \mathbb{E}^\lambda[f]$$

and

$$s \geq \inf_{\lambda \in \Xi_A^y(f)} \mathbb{E}^\lambda[f]$$

and hence for any $\epsilon > 0$ there exists $\mu_\epsilon \in \Xi_A^x(f)$ and $\nu_\epsilon \in \Xi_A^y(f)$ such that $t + \epsilon \geq \mathbb{E}^{\mu_\epsilon}[f]$ and $s + \epsilon \geq \mathbb{E}^{\nu_\epsilon}[f]$. For any $\alpha, \beta > 0$ with $\alpha + \beta = 1$ we have

$$\alpha t + \beta s + \epsilon \geq \mathbb{E}^{\alpha\mu_\epsilon + \beta\nu_\epsilon}[f].$$

Letting $\epsilon \to 0$ and noticing that $\alpha\mu_\epsilon + \beta\nu_\epsilon \in \Xi_A^{\alpha x + \beta y}(f)$ we obtain $\alpha(x,t) + \beta(y,s) \in \operatorname{epi}(H)$. It follows that H is convex and since for any $x \in A$

$$H(x) = \inf_{\lambda \in \Xi_A^x(f)} \mathbb{E}^\lambda[f] \leq \mathbb{E}^{\delta_x}[f] = f(x)$$

we have $H(x) \leq \overline{f}_A(x)$. Set $A_f = A \cap \{f < \infty\}$. For $(x_i)_{i=1}^n$ in A_f we have for all $t_i \geq \overline{f}_A(x_i)$ that $\sum_i \alpha_i(x_i, t_i) \in \operatorname{epi}(\overline{f}_A)$ with $\alpha_i > 0$ and $\sum_i \alpha_i = 1$ and hence that

$$\sum \alpha_i t_i \geq \overline{f}_A(\sum_i \alpha_i x_i).$$

Since this holds for all $t_i \geq \overline{f}_A(x_i)$ we obtain

$$\sum \alpha_i f(x_i) \geq \overline{f}_A(\sum_i \alpha_i x_i).$$

We now show that $\operatorname{epi}(H) \subset \operatorname{epi}(\overline{f}_A)$ from which it follows that $H(x) \geq \overline{f}_A(x)$. Assume $(x,t) \in \operatorname{epi}(H)$, then by the above argument for any $\epsilon > 0$ there exists $\mu_\epsilon \in \Xi_A^x(f)$ such that $t + \epsilon \geq \mathbb{E}^{\mu_\epsilon}[f]$. Since μ_ϵ is of finite support, there exist $(x_i)_i$ in A such that

$\mu_\epsilon = \sum_i \alpha_i \delta_{x_i}$. Hence we have

$$t + \epsilon \geq \mathbb{E}^{\mu_\epsilon}[f] = \sum_i \alpha_i f(x_i) \geq \sum_i \alpha_i \overline{f}_A(x_i) \geq \overline{f}_A(\sum_i \alpha_i x_i) = \overline{f}_A(x).$$

Since ϵ was arbitrary we obtain $(x,t) \in \mathrm{epi}(\overline{f}_A)$. □

Lemma 5.4.4. *For $f \leq g$ and $A \supset B$ we have*

$$\overline{f}_A \leq \overline{g}_B.$$

Proof. We have by Proposition 5.4.3 for any $x \in X$

$$\overline{f}_A(x) = \inf_{\lambda \in \Xi^x_A(f)} \mathbb{E}^\lambda[f] \leq \inf_{\lambda \in \Xi^x_B(f)} \mathbb{E}^\lambda[f] \leq \inf_{\lambda \in \Xi^x_B(f)} \mathbb{E}^\lambda[g] = \overline{g}_A(x).$$

□

The next Lemma shows that in order to calculate the infimum over all measures in $\Xi^x_A(f)$ it sufficies to consider those, whose support consists of less than or equal to $d+1$ points. Recall here that $A \subset X = \mathbb{R}^d$.

Lemma 5.4.5. *Set $\Xi^x_A(f,n) := \{\lambda \in \Xi^x_A(f) : \ |\operatorname{supp}(\lambda)| \leq n\}$. Then for any $f : X \to \overline{\mathbb{R}}$ we have*

$$\inf_{\Xi^x_A(f)} \mathbb{E}^\lambda[f] = \inf_{\Xi^x_A(f,d+1)} \mathbb{E}^\lambda[f]$$

Proof. Clearly

$$\inf_{\Xi^x_A(f)} \mathbb{E}^\lambda[f] \leq \inf_{\Xi^x_A(f,d+1)} \mathbb{E}^\lambda[f].$$

Set $A_f = A \cap \{f < \infty\}$ and fix $x \in \mathrm{conv}(A_f)$. Assume for the sake of contradiction that there is a minimal $n > d+1$ such that there exists a measure μ with $|\operatorname{supp}(\mu)| = n$ with

$$\mathbb{E}^\mu[f] < \inf_{\Xi^x_A(f,d+1)} \mathbb{E}^\lambda[f].$$

Since $\mu \in \Xi^x_A(f,n)$, there exist $x_1, \ldots, x_n \in A_f$ and $\alpha_1, \ldots, \alpha_n > 0$ with $\sum \alpha_i = 1$ such that $x = \sum_i \alpha_i x_i$ and $\mu = \sum \alpha_i \delta_{x_i}$. Consider the set $\mathrm{conv}(A_f) \times \{1\} \subset \mathbb{R}^{d+1}$. We have that $(x,1)$ is in the linear hull of $\{(x_i,1) : \ 1 \leq i \leq n\}$. Since $n > d+1$ these vectors are linearly dependent and hence there exist $\beta_1, \ldots, \beta_n$ not all zero such that $\sum \beta_i(x_i,1) = 0$. Since in particular this implies $\sum \beta_i = 0$ there must exist k, j with $\beta_j > 0$ and $\beta_k < 0$. Set $\gamma := \sup\{t : \ \alpha_i - t\beta_i \geq 0\} > 0$ and $\eta := \sup\{t : \alpha_i + t\beta_i \geq 0\} > 0$. Set

$$\mu_1 = \sum (\alpha_i - \gamma\beta_i)\delta_{x_i}$$

and

$$\mu_2 = \sum (\alpha_i + \eta\beta_i)\delta_{x_i}.$$

Since there exist i, j such that $\alpha_i - \gamma\beta_i = 0$ and $\alpha_j + \eta\beta_j = 0$ we have $|\operatorname{supp}(\mu_1)| < n$ and $|\operatorname{supp}(\mu_2)| < n$. By construction of μ_1, μ_2 we obtain the identity

$$\mu = \frac{\eta}{\gamma + \eta}\mu_1 + \frac{\gamma}{\gamma + \eta}\mu_2.$$

It follows that

$$\mathbb{E}^{\mu}[f] \geq \min\{\mathbb{E}^{\mu_1}[f], \mathbb{E}^{\mu_2}[f]\}.$$

But this contradicts the minimality of n and hence such an n cannot exist. □

Let Y be a finite dimensional real vector space.

Definition 5.4.6. *For ease of notation define for $P \subset Y \times X$ and $f : Y \times X \to \overline{\mathbb{R}}$ the map $H_P f : Y \times X \to \overline{\mathbb{R}}$ by*

$$H_P f(y, x) = \overline{f(y, \cdot)}_{P_y}(x).$$

It follows by Proposition 5.4.3 that $H_P f(y, \cdot)$ is convex.

Lemma 5.4.7. *For $f, g : Y \times X \to \overline{\mathbb{R}}$ with $g \leq f$ and $P_1 \subset P_2 \subset Y \times X$. Then*

$$H_{P_1} f \geq H_{P_2} g.$$

Proof. We have

$$\begin{aligned} H_{P_2} g(y, x) &= \overline{g(y, \cdot)}_{(P_2)_y}(x) \\ &\leq \overline{f(y, \cdot)}_{(P_1)_y}(x) = H_{P_1} f(y, x) \end{aligned}$$

where we have applied Lemma 5.4.4. □

Lemma 5.4.8. *Let $(P_n)_n$ be an increasing sequence of subsets of $Y \times X$ with $\cup P_n = P$ and $(f_n)_n$ a pointwise decreasing sequence of extended real functions on $Y \times X$ converging pointwise to f. Then we have*

$$\lim_{n\to\infty} H_{P_n} f_n(y, x) = H_P f(y, x)$$

pointwise for all $(y, x) \in Y \times X$.

Proof. By Proposition 5.4.3 for any $\epsilon > 0$ and $(y, x) \in Y \times X$ there exists a measure $\lambda \in \Xi^x_{P_y}(f(y, \cdot))$ such that

$$\mathbb{E}^{\lambda}[f(y, \cdot)] < H_P f(y, x) + \epsilon.$$

Since f_n converges pointwise to f, $\cup P_n = P$ and λ has finite support, there exists n_0 such that for all $n \geq n_0$ we have $\lambda \in \Xi^x_{(P_n)_y}(f(y, \cdot))$ and for all $z \in \operatorname{supp}(\lambda)$ we have $f_n(y, z) - f(y, z) \leq \epsilon$. This hold since $|\operatorname{supp}(\lambda)| < \infty$ and hence there must exists an n_0 such that $\operatorname{supp}(\lambda) \subset (P_{n_0})_y$. Thus for $n \geq n_0$

$$H_{P_n} f_n(y, x) \leq \mathbb{E}^{\lambda}[f_n(y, \cdot)] \leq \mathbb{E}^{\lambda}[f(y, \cdot)] + \epsilon < H_P f(y, x) + 2\epsilon.$$

Letting $\epsilon \to 0$ we obtain

$$\lim_n H_P f_n(y,x) \le H_P f(y,x).$$

Conversely since H is monotone in both it arguments P and f by Lemma 5.4.7 the reverse inequality is clear. $\square$

Definition 5.4.9. *Let $h : X^n \to \overline{\mathbb{R}}$. Consider h as a function on $X^{n-1} \times X$, then define $S_P^n h : X^{n-1} \to \overline{\mathbb{R}}$ by*

$$S_P^n h(y) = H_P h(y, p_{n-1}(y))$$

where $y \in X^{n-1}$ and $p_{n-1} : X^{n-1} \to X$ the projection onto the last component. Set $P_k := (p_1, \dots, p_k)(P)$ and for $f : X \to \overline{\mathbb{R}}$ define

$$S_P f = S_{P_2}^2 \circ \dots \circ S_{P_{n-1}}^{n-1} \circ S_P^n (f \circ p_n).$$

and for $h : Z \to \overline{\mathbb{R}}$ set

$$\widetilde{S}_P h = S_{P_2}^2 \circ \dots \circ S_{P_{n-1}}^{n-1} \circ S_P^n (h)$$

Remark 5.4.10. *This definition for S_P is equivalent to Definition 5.2.4. This can be seen from*

$$\begin{aligned} H_P h(x, p_{n-1}(x)) &= \overline{h(x,\cdot)}_{P_x}(p_{n-1}(x)) \\ &= \sup\{g(p_{n-1}(x)) : \ g \text{ convex and } g(y) \le h(x,y) \text{ for } y \in P_x\}. \end{aligned}$$

Definition 5.4.11. *Define for $h : X^n \to \overline{\mathbb{R}}$ and $\mu \in M^1(X)$ with finite support the set*

$\mathcal{D}_P^\mu(f) :=$
$\{\lambda \in M(X^n) : \lambda$ *martingale measure with* $\operatorname{supp}(\lambda) < \infty,\ (p_1)_*(\lambda) = \mu$ *and* $\lambda(P \cap \{f < \infty\}) = 1\}$.

The next duality result will be the crucial ingredient to prove most of the existence results on martingales with restricted support.

Proposition 5.4.12. *Let $P \subset X^n$, $f : X^n \to \overline{\mathbb{R}}$ and $\mathcal{D}_P^\mu(f)$ as in Definition 5.4.11. We have*

$$\inf_{\lambda \in \mathcal{D}_P^{\delta_{x_0}}(h)} \mathbb{E}^\lambda[h] = \widetilde{S}_P h(x_0). \tag{5.4.1}$$

for $x_0 \in X$. In particular for $f : X \to \mathbb{R}$ we obtain

$$\inf_{\lambda \in \mathcal{D}_P^{\delta_{x_0}}(f \circ p_n)} \mathbb{E}^\lambda[f \circ p_n] = S_P f(x_0). \tag{5.4.2}$$

Proof. We first assume $n = 2$. For $h : X^2 \to \overline{\mathbb{R}}$, set $h_x := h(x, \cdot)$. By Proposition 5.4.3 we have

$$\inf_{\lambda \in \mathcal{D}_P^{\delta_{x_0}}(h)} \mathbb{E}^\lambda[h] = \inf_{\lambda \in \Xi_{P_{x_0}}^{x_0}(h_{x_0})} \mathbb{E}^\lambda[h_{x_0}] = H_P h(x_0, x_0) = \widetilde{S}_P h(x_0).$$

For $n > 2$ we apply induction. Consider $X^n = X^{n-1} \times X$ and for $h : X^{n-1} \times X \to \overline{\mathbb{R}}$ again set $h_y = h(y, \cdot)$ with $y \in X^{n-1}$. Any $\lambda \in \mathcal{D}_P^{\delta_{x_0}}(h)$ has a disintegration $\lambda(dy, dx) = \eta(y, dx) * \mu(dy)$ over $X^{n-1} \times X$ with $y \in X^{n-1}$ and $x \in X$, where $\mu = (p_1, \dots, p_{n-1})_* \lambda$ is a martingale measure and η is a Markov kernel with finite support because λ has finite support. Set $p := (p_1, \dots, p_{n-1})$ the projection onto X^{n-1}. We obtain

$$\begin{aligned}
\inf_{\lambda \in \mathcal{D}_P^{\delta_{x_0}}(h)} \mathbb{E}^\lambda[h] &= \inf_{\lambda \in \mathcal{D}_P^{\delta_{x_0}}(h)} \sum_{y \in \mathrm{supp}(p_* \lambda)} \lambda(\{y\}) \inf_{\eta \in \Xi_{P_y}^{p_{n-1}(y)}(h_y)} \mathbb{E}^\eta[h_y] \\
&= \inf_{\lambda \in \mathcal{D}_P^{\delta_{x_0}}(h)} \sum_{y \in \mathrm{supp}(p_* \lambda)} \lambda(\{y\}) H_P h(y, p_{n-1}(y)) \\
&= \inf_{\lambda \in \mathcal{D}_P^{\delta_{x_0}}(h)} \mathbb{E}^{p_* \lambda}[S_P^n h].
\end{aligned}$$

It remains to show that $p_* \mathcal{D}_P^{\delta_{x_0}}(h) = \mathcal{D}_{P_{n-1}}^{\delta_{x_0}}(S_P^n h)$. Assume $\mathcal{D}_P^{\delta_{x_0}}(h) \neq \emptyset$. Clearly for any $\lambda \in \mathcal{D}_P^{\delta_{x_0}}(h)$, $p_* \lambda$ is a martingale measure on P_{n-1} with finite support. Write $\lambda(dy, dx) = \eta(y, dx) p_* \lambda(dy)$. Since $\lambda(\{h < \infty\}) = 1$ for all $y \in \mathrm{supp}(p_* \lambda)$ we have $\eta(y, \cdot) \in \Xi_{P_y}^{p_{n-1}(y)}(h_y)$. Thus $S_P^n h(y) < \infty$ and hence $p_* \lambda \in \mathcal{D}_{P_{n-1}}^{\delta_{x_0}}(S_P^n h)$. Conversely for $\mu \in \mathcal{D}_{P_{n-1}}^{\delta_{x_0}}(S_P^n h)$ we have for any $y \in \mathrm{supp}(\mu)$ that $S_P^n h < \infty$ and hence by Proposition 5.4.3

$$S_P^n h(y) = \inf_{\eta \in \Xi_{P_y}^{p_{n-1}(y)}(h_y)} \mathbb{E}^\eta[h_y]$$

and thus the set $\Xi_{P_y}^{p_{n-1}(y)}(h_y) \neq \emptyset$. Choose a transition kernel η such that for $y \in \mathrm{supp}(\mu)$ we have $\eta(y, \cdot) \in \Xi_{P_y}^{p_{n-1}(y)}(h_y)$. It follows that $\lambda(dy, dx) := \eta(y, dx)\mu(dy)$ is in $\mathcal{D}_P^{\delta_{x_0}}(h)$ and satisfies $p_* \lambda = \mu$. Hence

$$\inf_{\lambda \in \mathcal{D}_P^{\delta_{x_0}}(h)} \mathbb{E}^{p_* \lambda}[S_P^n h] = \inf_{\lambda \in \mathcal{D}_{P_{n-1}}^{\delta_{x_0}}(S_P^n h)} \mathbb{E}^\lambda[S_P^n h].$$

This combined with the induction hypothesis yields

$$\inf_{\lambda \in \mathcal{D}_{P_{n-1}}^{\delta_{x_0}}(S_P^n h)} \mathbb{E}^\lambda[S_P^n h] = \widetilde{S}_{P_{n-1}}(S_P^n h)(x) = S_{P_2}^2 \circ \cdots \circ S_{P_{n-1}}^{n-1} \circ S_P^n h(x) = \widetilde{S}_P h(x).$$

Applying (5.4.1) to $h = f \circ p_n$ for $f : X \to \overline{\mathbb{R}}$ yields (5.4.2). □

Corollary 5.4.13. *Denote by $\mathcal{D}_P^\mu(f, d+1) \subset \mathcal{D}_P^\mu(f)$ the set of martingale measures such that each transition probability measure is supported on at most $d+1$ points, i.e. for any $\mathbb{Q} \in \mathcal{D}_P^\mu(f, d+1)$ we have that $\mathbb{Q}(dx_i | x_1, \dots, x_{i-1})$ is supported on at most $d+1$ distinct points. Then*

$$\inf_{\lambda \in \mathcal{D}_P^{\delta_x}(f, d+1)} \mathbb{E}^\lambda[f] = S_P f(x).$$

Proposition 5.4.12 and Corollary 5.4.13 provide a numerical way to calculate $S_P f$

by taking the infimum over discrete finite support martingales, where in each step the martingale can only move to at most $d+1$ points from a given point. Clearly by Lemma 5.4.5 we obtain the following corollary.

5.4.2. A Strassen-type result on martingales with restricted support

In this subsection, we first demonstrate that S_P is the appropriate operator for studying martingales supported on P. This analysis will reveal why the conditions in Theorem 5.2.7 are necessary. To establish sufficiency, we then use a functional analytic approach to show that understanding discrete martingales supported on P suffices to charaterize all martingales supported on P.

Definition 5.4.14. *Let Ω be a measurable space and $f : \Omega \mapsto [-\infty, \infty]$ be measurable. We say that f is $\mathbb{Q}$-integrable for some probability measure $\mathbb{Q}$ if $\int_{\{f>0\}} f d\mathbb{Q} < \infty$.*

We introduce this notion of integrability to mainly avoid issues when f is not integrable in the classical sense.
By Theorem 5.B.8 and Theorem 5.B.9 for σ-compact $P \subset Z$ for $h \in C_{\psi_Z}(Z)$ and $f \in C_\psi(X)$ the functions $H_P h$ and $S_P f$ are Borel measurable. Thus given suitable integrability conditions on $H_P h$ and $S_P f$ their integral with respect to a finite Borel measure is well defined. For the remainder of this section, we will assume this measurability without further mention.
The next Theorem shows a Jensen type inequality that any martingale supported on P must satisfy.

Theorem 5.4.15. *Let $P \subset X^n$ be σ-compact and let $\mathbb{Q} \in \mathcal{M}(P)$, i.e. a martingale measure with $\mathbb{Q}(P) = 1$. Let $f : X^n \to \overline{\mathbb{R}}$ be σ-lower semi-continuous. If f is $\mathbb{Q}$-integrable then $(S^n_P f) \circ (p_1, \dots, p_{n-1})$ is $\mathbb{Q}$-integrable and*

$$\mathbb{E}^{\mathbb{Q}}[f] \geq \mathbb{E}^{\mathbb{Q}}[(S^n_P f) \circ (p_1, \dots, p_{n-1})].$$

This further implies that for $f : X \to \overline{\mathbb{R}}$, $S_P f$ is μ-integrable and

$$\mathbb{E}^{\mathbb{Q}}[f \circ p_n] \geq \mathbb{E}^{\mu}[S_P f]$$

where $\mu = (p_1)_*(\mathbb{Q})$.

Proof. Consider $X^n = X^{n-1} \times X$ and let

$$\mathbb{Q}(dy, dx) = \eta(y, dx)\nu(dy)$$

be the disintegration of $\mathbb{Q}$, where $y \in X^{n-1}$ and $x \in X$. Set $g = \mathbb{1}_{\{S^n_P f>0\}} \circ (p_1, \dots, p_{n-1})$ and $A := \{S^n_P f > 0\}$. Hence

$$\infty > \mathbb{E}^{\mathbb{Q}}[gf] = \int \mathbb{1}_A(y) \int f(y, x)\eta(y, dx)\nu(dy)$$

since $gf\mathbb{1}_{\{f>0\}} \leq f\mathbb{1}_{\{f>0\}}$ and f is $\mathbb{Q}$-integrable. We obtain

$$\int f(y,x)\eta(y,dx) \geq \int H_P f(y,x)\eta(y,dx) \geq H_P(y,\pi(\eta(y,\cdot))) = H_P f(y,p_{n-1}(y)) = S_P^n f(y)$$

ν-almost surely, where $\pi(\eta(y,\cdot))$ denotes the barycenter of $\eta(y,\cdot)$. Indeed, the first inequality above follows from the fact that $f(y,x) \geq H_P f(y,x)$ $\eta(y,\cdot)$-almost surely since $\eta(y,P_y) = 1$ and the second from Jensen's inequality as stated in Lemma 5.C.3. Hence we calculate

$$\begin{aligned}\infty > \mathbb{E}^{\mathbb{Q}}[gf] &= \int \mathbb{1}_A(y) \int f(y,x)\eta(y,dx)\nu(dy) \\ &\geq \int \mathbb{1}_A(y) S_P^n f(y)\nu(dy) = \mathbb{E}^{\mathbb{Q}}[(\mathbb{1}_{\{S_P^n f>0\}} S_P^n f)\circ(p_1,\dots,p_{n-1})].\end{aligned}$$

This shows that $(S_P^n f)\circ(p_1,\dots,p_{n-1})$ is $\mathbb{Q}$-integrable. Repeating the same calculation without g yields

$$\mathbb{E}^{\mathbb{Q}}[f] = \int\int f(y,x)\eta(y,dx)\nu(dy) \geq \int S_P^n f(y)\nu(dy) = \mathbb{E}^{\mathbb{Q}}[(S_P^n f)\circ(p_1,\dots,p_{n-1})].$$

Applying this inequality successively yields the second statement. □

Lemma 5.4.16. *Let $P \subset Y \times X$, then H_P is superlinear (see Definition 5.A.8) on the space of functions $f : Y \times X \to \overline{\mathbb{R}}$.*

Proof. Let $f, g : Y \times X \to \overline{\mathbb{R}}$, then

$$\begin{aligned}&H_P(f+g)(y,x) = \overline{(f+g)(y,\cdot)}_{P_y}(x) \\ &= \sup\{h(x) :\ h \text{ convex and } h(x) \leq f(y,x) + g(y,x),\ (y,x) \in P\} \\ &\geq \sup\{a(x) + b(x) :\ a, b \text{ convex and } a(x) \leq f(y,x) \text{ and } b(x) \leq g(y,x),\ (y,x) \in P\} \\ &= \overline{f(y,\cdot)}_{P_y}(x) + \overline{g(y,\cdot)}_{P_y}(x) = H_P f(y,x) + H_P g(y,x).\end{aligned}$$

The positive homogeneity is obvious since the positive multiple of a convex function is convex as well. □

Lemma 5.4.17. *Consider $X^n = X^{n-1} \times X$ and let $g : X^{n-1} \times X \to \overline{\mathbb{R}}$ be convex. Assume $f \geq g$, then for all $y \in X^{n-1}$ we have $S_P^n f(y) \geq g(y, p_{n-1}(y))$. In particular for $f, g : X \to \overline{\mathbb{R}}$ with $f \geq g$ and g convex we have*

$$S_P f(x) \geq g(x)$$

for all $x \in X$.

Proof. Clearly $g(y,\cdot)$ is convex and $g(y,x) \leq f(y,x)$ for all y, x and hence in particular for $(y,x) \in P$. Thus

$$\begin{aligned} S_P^n f(y) &= H_P f(y, p_{n-1}(y)) \\ &= \sup\{h(p_{n-1}(y)) : \underbrace{h \text{ convex and } h(x) \le f(y,x) \text{ for } (y,x) \in P}_{\ni g(y,\cdot)}\} \ge g(y, p_{n-1}(y)). \end{aligned}$$

For the second statement apply the above successively to $f \circ p_n$ and $g \circ p_n$. $\square$

Assumption 5.4.18. *Set $\psi(x) = \frac{1}{1+||x||^t}$ for $x \in X$ and $t \ge 1$ and for $Z = X^n$ define $\psi_Z : Z \to \mathbb{R}$ via*

$$\frac{1}{\psi_Z} = \sum_{i=1}^n \frac{1}{\psi \circ p_i}.$$

For the next lemma recall the notations introduced in Definition 5.2.6.

Lemma 5.4.19. *For every closed $P \subset Z$ and $\nu \in M_\psi^1(X)$ the sets $\mathcal{M}_\psi(P)$ and ${}^\nu\mathcal{M}_\psi(P)$ of martingale measures are convex and closed in the weak-$*$-topology induced by $C_{\psi_Z}(Z)$.*

Proof. $\mathbb{Q}$ is a martingale measure in Z with $\mathbb{Q}(P) = 1$ if and only if for all $1 < i \le n$ and for all $f \in C_b(X^{i-1})$ we have

$$\mathbb{E}^{\mathbb{Q}}[(f \circ p_{1,i-1})(p_i - p_{i-1})] = 0,$$

and for all $g \in C_{\psi_Z}(Z)$ with $P \subset g^{-1}(\{0\})$ we have

$$\mathbb{E}^{\mathbb{Q}}[g] = 0.$$

Hence the set of martingale measure $\mathcal{M}_\psi(P)$ is closed and convex since the functions $(f \circ p_{1,i-1})(p_i - p_{i-1})$ and g are in $C_{\psi_Z}(Z)$. Set

$$A := \{\mathbb{Q} \in M_{\psi_Z}(Z) | \forall h \in C_\psi(X) : \mathbb{E}^{\mathbb{Q}}[h \circ p_n] - \mathbb{E}^\nu[h] = 0\}.$$

Since $h \circ p_n \in C_{\psi_Z}(Z)$, A is closed and convex and thus

$${}^\nu\mathcal{M}_\psi(P) = A \cap \mathcal{M}_\psi(P)$$

is closed and convex. $\square$

Lemma 5.4.20. *For $\nu \in M_\psi^1(X)$ the set ${}^\nu\mathcal{M}_\psi(Z)$ is weak-$*$-compact with respect to $C_{\psi_Z}(Z)$.*

Proof. By Lemma 5.4.19 ${}^\nu\mathcal{M}_\psi(Z)$ is closed and hence we only need to show that it is relatively compact. The map $\phi_Z : M_{\psi_Z}(Z) \to M(Z)$ given by $\phi(\mathbb{Q})(E) = \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_E \frac{1}{\psi_Z}]$ for Borel measurable sets $E \subset Z$, is a homemorphism, where $M(Z)$ is equipped with the usual weak convergence. Hence it is enough to show that the set $\phi_Z({}^\nu\mathcal{M}_\psi(Z))$ is tight and hence compact in $M(Z)$. For $\mathbb{Q} \in {}^\nu\mathcal{M}_\psi(Z)$ set $q_m = (p_m)_*(\mathbb{Q})$ where $1 \le m \le n$.

By definition of $M_{\psi_Z}(Z)$ we have $q_m \in M_\psi(X)$.
For $k \geq 1$ set

$$r_k(x) = \begin{cases} 0 & \text{if } x \in B_{k^t-1} \\ 1 + ||x||^t - k^t & \text{else} \end{cases}$$

where $B_{k^t}(0) \subset X$ is the Ball at 0 with radius k^t and t is the exponent in ψ. Clearly $r_k(x) \in C_\psi(X)$ since $|r_k\psi| \leq 1$. We have $(1 + ||x||^t)\mathbb{1}_{X\setminus B_{2k^t-1}}(x) \leq 2r_k(x)$ since for $x \in X \setminus B_{2k^t-1}$ we have

$$\frac{1 + ||x||^t}{1 + ||x||^t - k^t} \leq \frac{2k^t}{k^t} = 2.$$

Since r_k is convex and $(q_m)_m$ are in convex order we obtain from Jensen's inequality for all q_m

$$\mathbb{E}^{q_m}[\frac{1}{\psi}\mathbb{1}_{X\setminus B_{2k^t-1}}] = \mathbb{E}^{q_m}[(1 + ||x||^t)\mathbb{1}_{X\setminus B_{2k^t-1}}(x)] \leq 2\mathbb{E}^{q_m}[r_k] \leq 2\mathbb{E}^{q_n}[r_k] = 2\mathbb{E}^\nu[r_k].$$

Set $A_k = Z \setminus B_{2k^t-1} \times \cdots \times B_{2k^t-1}$, then we have

$$\begin{aligned} \phi_Z(\mathbb{Q})(A_k) &= \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{A_k}\frac{1}{\psi_Z}] = \sum_{i=1}^n \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{A_k}\frac{1}{\psi} \circ p_i] \leq \sum_{i=1}^n \mathbb{E}^{\mathbb{Q}}[(\mathbb{1}_{X\setminus B_{2k^t-1}}\frac{1}{\psi}) \circ p_i] \\ &\leq \sum_{i=1}^n \mathbb{E}^{q_i}[\mathbb{1}_{X\setminus B_{2k^t-1}}\frac{1}{\psi}] \leq 2n\mathbb{E}^\nu[r_k] \leq 2n\mathbb{E}^\nu[\mathbb{1}_{X\setminus B_{k^t-1}}\frac{1}{\psi}] \underset{k\to\infty}{\to} 0. \end{aligned} \tag{5.4.3}$$

Since $B_{2k^t-1} \times \cdots \times B_{2k^t-1}$ is compact and $\mathbb{Q} \in {}^\nu\mathcal{M}_\psi(Z)$ was arbitrary it follows that ${}^\nu\mathcal{M}_\psi(Z)$ is weak-$*$- compact. $\square$

Theorem 5.4.21. *Let Λ be a non empty closed convex subset of $\mathcal{M}_\psi(Z)$ of martingale measures and let μ, ν be a probability measures in $M^1_\psi(X)$, then there exists $\mathbb{Q} \in \Lambda$ with $(p_n)_*\mathbb{Q} = \nu$ and $(p_1)_*\mathbb{Q} = \mu$ if and only if for any $f, g \in C_\psi(X)$ we have*

$$\mathbb{E}^\nu[f] + \mathbb{E}^\mu[g] \geq \inf_{\lambda\in\Lambda} \mathbb{E}^\lambda[f \circ p_n + g \circ p_1] \tag{5.4.4}$$

Proof. We first show that the inequality condition (5.4.4) implies the existence of a martingale measure $\mathbb{Q}$ in Λ with final marginal ν. The map $\phi : M_\psi(X) \to M(X)$ given by

$$\mu \mapsto (\phi(\mu) : E \mapsto \mathbb{E}^\mu[\mathbb{1}_E\frac{1}{\psi}])$$

for Borel measurable sets E, is a homeomorphism and for any $\mu \in M^1_\psi(X)$, $\phi(\mu)$ is a positive finite Borel measure. By Theorem 8.3.2 in [17] it follows that the set $\phi(M^1_\psi(X))$ equipped with the weak convergence is metrizable. Hence $M^1_\psi(X)$ is metrizable in the weak-$*$- topology induced by $C_\psi(X)$. The same argument applied to Z shows that $M^1_{\psi_Z}(Z)$ is metrizable in the weak-$*$ topology induced by $C_{\psi_Z}(Z)$. Equip $M_\psi(X) \times M_\psi(X)$ with the product topology. Set

$$F = \{((p_1)_*, (p_n)_*)(\mathbb{Q}) | \mathbb{Q} \in \Lambda\} \subset M^1_\psi(X) \times M^1_\psi(X).$$

Clearly F is convex because Λ is convex. Denote by $\mathrm{cl}(F)$ the closure of F. If $(\mu, \nu) \notin \mathrm{cl}(F)$ then by the Hahn-Banach separation theorem there exists some $f, g \in C_\psi(X)$ such that

$$\int_X f d\nu + \int_X g d\mu < \inf_{\lambda \in \Lambda} \int_Z [f \circ p_n + g \circ p_1] d\lambda.$$

Hence we have $(\mu, \nu) \in \mathrm{cl}(F)$ and thus there exists a sequence $(\mu_i, \nu_i)_i$ in F such that $(\mu_i, \nu_i) \to (\mu, \nu)$ in the weak-$*$- topology induced by $C_\psi(X) \times C_\psi(X)$. In particular $\nu_i \to \nu$ in the weak-$*$- topology induced by $C_\psi(X)$. It follows that the family $(\nu_i)_i$ is weak-$*$-compact and hence $(\phi(\nu_i))_i$ is tight, i.e.,

$$\lim_{k \to \infty} \sup_i \mathbb{E}^{\nu_i}[\mathbb{1}_{X \setminus B_{2k^t - 1}} \frac{1}{\psi}] = 0$$

where B_{2k^t-1} is the closed ball with radius $2k^t - 1$ and center 0. Since each $(\mu_i, \nu_i) \in F$ there exists $\mathbb{Q}_i$ such that $(p_n)_*(\mathbb{Q}_i) = \nu_i$. By (5.4.3)

$$\lim_{k \to \infty} \sup_i \mathbb{E}^{\mathbb{Q}_i}[\mathbb{1}_{A_k} \frac{1}{\psi_Z}] = 0$$

and hence the family $(\mathbb{Q}_i)_i$ is weak-$*$-compact in $M_{\psi_Z}(Z)$. Let $\mathbb{Q}$ be a cluster point of $(\mathbb{Q}_i)_i$, then $\mathbb{Q} \in \Lambda$ since Λ is closed, $(p_1)_*(\mathbb{Q}) = \mu$ and $(p_n)_*(\mathbb{Q}) = \nu$ since for any $f \in C_\psi(X)$, $f \circ p_1$ and $f \circ p_n$ are in $C_{\psi_Z}(Z)$. Conversely if there exists such a martingale measure $\mathbb{Q} \in \Lambda$ then the inequality is trivially satisfied. □

Corollary 5.4.22. *Let $P \subset Z$ be closed. The sets ${}^\nu\mathcal{M}_\psi(P)$ and ${}^\nu_\mu\mathcal{M}_\psi(P)$ are weak-$*$-compact with respect to $C_{\psi_Z}(Z)$.*

Proof. We have

$${}^\nu\mathcal{M}_\psi(P) = {}^\nu\mathcal{M}_\psi(Z) \cap {}^\nu\mathcal{M}_\psi(P)$$

and thus by Lemma 5.4.19 and Lemma 5.4.20 is closed and compact. Denote by $I(\mu) \subset M_{\psi_Z}(Z)$ the subset of finite signed measures with initial distribution μ. $I(\mu)$ is closed since it is the intersection of kernels of the continuous functionals given by

$$\lambda \mapsto \mathbb{E}^\lambda[f \circ p_1]$$

where $f \in C_\psi(X)$. We have

$${}^\nu_\mu\mathcal{M}_\psi(Z) = I(\mu) \cap {}^\nu\mathcal{M}_\psi(P)$$

and thus it is closed and compact. □

Theorem 5.4.23. *Let μ and ν be probability measures in $M_\psi(X)$ and let $P \subset Z$ be closed. Then there exists a martingale measure $\mathbb{Q} \in {}_\mu^\nu\mathcal{M}_\psi(P)$ if and only if for all $f \in C_\psi(X)$, $S_P f$ is μ-integrable in the sense of Definition 5.4.14 and we have*

$$\mathbb{E}^\nu[f] \geq \mathbb{E}^\mu[S_P f]. \tag{5.4.5}$$

Proof. We first show that the inequality (5.4.5) being satisfied for all $f \in C_\psi(X)$ implies the non emptyness of ${}_\mu^\nu\mathcal{M}_\psi(P)$ and the μ-integrability of $S_P f$. In order to apply Theorem 5.4.21 for

$$\Lambda := \{\mathbb{Q} : \ \mathbb{Q} \text{ martingale measure and } \mathbb{Q}(P) = 1\} = \mathcal{M}_\psi(P)$$

we need to show that it is non empty and closed. Since P is closed, by Lemma 5.4.19 Λ is closed. Suppose by contradiction that it is empty, then $\mathcal{D}_P^{\delta_x}(f)$ (see Definition 5.4.11) is empty for every $f \in C_\psi(X)$ and hence $S_P f \equiv \infty$ by Proposition 5.4.12. But this contradicts $S_P f$ being μ-integrable. Hence Λ is non empty and by Lemma 5.4.19 it is closed. For $f, g \in C_\psi(X)$ we have

$$\mathbb{E}^\nu[f] + \mathbb{E}^\mu[g] \geq \mathbb{E}^\mu[S_P f + g] \geq \inf_{x \in X} S_P f(x) + g(x) = \inf_{x \in p_1(P)} S_P f(x) + g(x) \tag{5.4.6}$$

where the first inequality follows from (5.4.5) and the last equality follows from $S_P f(x) = \infty$ if $x \notin p_1(P)$ by Proposition 5.4.12. We have by Proposition 5.4.12

$$\inf_{\lambda \in \mathcal{D}_P^{\delta_x}(f \circ p_n)} \mathbb{E}^\lambda[f \circ p_n] = S_P f(x). \tag{5.4.7}$$

Hence combining (5.4.6) and (5.4.7) we obtain

$$\begin{aligned}
\inf_{\mathbb{Q} \in \Lambda} \mathbb{E}^\mathbb{Q}[f \circ p_n + g \circ p_1] &\leq \inf_{x \in p_1(P)} \inf_{\lambda \in \mathcal{D}_P^{\delta_x}(f \circ p_n)} \mathbb{E}^\lambda[f \circ p_n] + g(x) \\
&= \inf_{x \in p_1(P)} S_P f(x) + g(x) \leq \mathbb{E}^\nu[f] + \mathbb{E}^\mu[g],
\end{aligned}$$

which is exactly the condition of Theorem 5.4.21, implying the existence of $\mathbb{Q} \in \Lambda$ with $(p_1)_*(\mathbb{Q}) = \mu$ and $(p_n)_*(\mathbb{Q}) = \nu$. Conversely since every closed set $P \subset Z$ is σ-compact by Theorem 5.4.15 we have

$$\mathbb{E}^\mu[S_P f] \leq \mathbb{E}^\nu[f].$$

□

5.4.3. Duality results

From a functional analytic point of view Theorem 5.4.23 hints at a deeper relation between S_P and the final distributions of martingale measures supported on P. Specifically that $f \mapsto \mathbb{E}^\mu[S_P f]$ acts as the support function for the set of final distributions of measures in ${}_\mu\mathcal{M}_\psi(P)$. In order to formalize this relationship, we need to define the appropriate topology on $M_\psi(X)$.

Definition 5.4.24.

- *Denote by β the strict topology on $C_b(X)$ generated by the seminorms* $\sup|f\phi|$ *where ϕ in $C_0(X)$, the space of continuous real functions vanishing at infinity.*

- *Consider the bijective map $\Psi : C_\psi(X) \to C_b(X) : f \mapsto f\psi$, where $C_b(X)$ is equipped with the strict topology β. Let β_ψ be the topology on $C_\psi(X)$ making Ψ a homeomorphism, i.e. the initial topology induced by Ψ.*

Remark 5.4.25. *By [20] the dual space of $(C_b(X), \beta)$ is the space $M(X)$ of all finite signed Borel measures on X.*

Proposition 5.4.26. *The dual space of $(C_\psi(X), \beta_\psi)$ is given by $M_\psi(X)$.*

Proof. Since $\Psi : C_\psi(X) \to C_b(X) : f \mapsto f\psi$ is a vector space homeomorphism, its dual map

$$\Psi^* : C_b(X)^* \to C_\psi(X)^* : (f \mapsto l(f)) \mapsto (f \mapsto l \circ \Psi(f) = l(f\psi))$$

is a bijection between $C_b(X)^*$ and $C_\psi(X)^*$. This induces a bijection $d\lambda \mapsto \psi d\lambda$ for $\lambda \in M(X)$. Hence $\psi d\lambda \in M_\psi(X)$ and $C_\psi(X)^* = M_\psi(X)$. □

Lemma 5.4.27. *For any $\nu \in M^1_\psi(X)$ there exists $\phi \in C_0(X)$ such that $\frac{1}{\psi\phi}$ is convex and*

$$\mathbb{E}^\nu\left[\frac{1}{\psi\phi}\right] < \infty$$

Proof. Since the class of functions with the single element $\frac{1}{\psi}$ is uniformly integrable, by the Vallée Poussin Theorem there exists a non-negative increasing convex function $\Phi : \mathbb{R} \to \mathbb{R}_{>0}$ such that $\Phi(\frac{1}{\psi})$ is integrable and

$$\lim_{|y|\to\infty} \frac{\Phi(y)}{|y|} = \infty.$$

Since $\frac{1}{\psi}$ is convex the composition $\Phi(\frac{1}{\psi})$ is convex. Set $\phi = \frac{1/\psi(x)}{\Phi(1/\psi(x))}$. Clearly $\phi \in C_0(X)$ and $\frac{1}{\phi\psi} = \Phi(\frac{1}{\psi})$. □

Theorem 5.4.28. *Let $\mu \in M^1_\psi(X)$, $P \subset Z$ be closed and assume that ${}_\mu\mathcal{M}_\psi(P) \neq \emptyset$. Set $\mathcal{N}_\mu(P) := (p_n)_* \, {}_\mu\mathcal{M}_\psi(P)$, i.e. the set of final marginal distribution of martingale measures in ${}_\mu\mathcal{M}_\psi(P)$. Then for any $f \in C_\psi(X)$ we have the following strong duality*

$$\inf_{\nu\in\mathcal{N}_\mu(P)} \mathbb{E}^\nu[f] = \mathbb{E}^\mu[S_P f].$$

Proof. Choose $\nu \in \mathcal{N}_\mu(P)$. By Lemma 5.4.27 there exists $\phi \in C_0(X)$ such that $\frac{1}{\psi\phi}$ has finite expectation under ν. We first show that the map $f \mapsto \mathbb{E}^\mu[S_P f]$ for $f \in C_\psi(X)$ is upper semi-continuous with respect to the topology β_ψ. In order to show upper semi-continuity we need to show that for any $f \in C_\psi(X)$ and $r \in \mathbb{R}$ with $\mathbb{E}^\mu[S_P f] < r$ there exists an open neighbourhood $\mathcal{U}$ of f in β_ψ such that

$$\sup_{g \in \mathcal{U}} \mathbb{E}^\mu[S_P g] < r.$$

Set $\mathcal{U}_\epsilon := \{g \in C_\psi(X) : |f - g|\psi\phi < \epsilon\}$. For every ϵ, $\mathcal{U}_\epsilon$ is open in the β_ψ topology. Clearly for every g in $\mathcal{U}_\epsilon$ we have the bound

$$g(x) \leq f(x) + \frac{\epsilon}{\psi(x)\phi(x)}.$$

By the choice of ϕ we have that $f + \frac{\epsilon}{\psi\phi}$ is ν-integrable and hence by Theorem 5.4.15 $S_P(f + \frac{\epsilon}{\psi\phi})$ is μ-integrable for all $\epsilon \geq 0$. Since S_P is order preserving (see Lemma 5.4.7 together with Definition 5.4.9) we have that $S_P(f + \frac{\epsilon}{\psi\phi})$ is a decreasing sequence of functions as ϵ decreases. Hence by monotone convergence and Lemma 5.4.8 we obtain

$$\lim_{\epsilon \to 0} \mathbb{E}^\mu[S_P(f + \frac{\epsilon}{\psi\phi})] = \mathbb{E}^\mu[\lim_{\epsilon \to 0} S_P(f + \frac{\epsilon}{\psi\phi})] = \mathbb{E}^\mu[S_P f] < r.$$

Thus there exists $\epsilon_0 > 0$ such that for all $\epsilon \leq \epsilon_0$, it holds that $\mathbb{E}^\mu[S_P(f + \frac{\epsilon}{\psi\phi})] < r$ and this implies that $\mathcal{U}_{\epsilon_0}$ is the desired open set. Hence, the map $f \mapsto \mathbb{E}^\mu[S_P f]$ is upper semi-continuous and by Lemma 5.A.8 also superlinear. Moreover, $\mathbb{E}^\mu[S_P f] < \infty$ since $S_P f$ is μ-integrable. It follows by Fenchel-Moreau (see Theorem 5.A.9) that $\mathbb{E}^\mu[S_P f]$ is the negative of a supporting function, i.e. there exists a closed convex set $A \subset C_\psi(X)^* = M_\psi(X)$ such that

$$\mathbb{E}^\mu[S_P f] = \inf_{h^* \in A} h^*(f).$$

We are left to show that the elements in A are probability measures and satisfy the conditions of Theorem 5.4.23. By Lemma 5.4.17 we have $S_P \mathbf{1} \geq \mathbf{1}$ and hence $\inf_{h^* \in A} h^*(\mathbf{1}) = \mathbb{E}^\mu[S_P \mathbf{1}] \geq 1$ where $\mathbf{1} : X \to \mathbb{R}$ is the constant unit valued function. Furthermore $-\sup_{h^* \in A} h^*(\mathbf{1}) = \inf_{h^* \in A} h^*(-\mathbf{1}) \geq -1$. This shows that every measure in A has mass 1. Assume there is a signed measure in A which is negative on some open set $B \subset X$. It follows that there exists non negative $f \in C_\psi(X)$ with $h^*(f) < 0$. Then we have

$$0 > \inf_{h^* \in A} h^*(f) = \mathbb{E}^\mu[S_P f] \geq \mathbb{E}^\mu[S_P \mathbf{0}] \geq \mathbb{E}^\mu[0] = 0$$

which is a contradiction. Hence every measure in A is non negative on all open sets and thus by outer regularity non negative. Since every $h^* \in A$ with $h^*(f) = \int f d\lambda$ satisfies $\mathbb{E}^\lambda[f] \geq \mathbb{E}^\mu[S_P f]$ it follows by Theorem 5.4.23 that $A \subset \mathcal{N}_\mu(P)$. By Theorem 5.4.15 we also have the reverse inclusion. □

Remark 5.4.29. *We make use of the term duality in the theorem above as $S_P f$ itself is the solution of an optimization problem, see Definition 5.2.4 and Example 3.*

Corollary 5.4.30. *For $P \subset Z$ closed and $\mu \in M^1_\psi(X)$ we have ${}_\mu\mathcal{M}_\psi \neq \emptyset$ if and only if $f \mapsto \mathbb{E}^\mu[S_P f]$ is upper semi-continuous and does not attain ∞ on $(C_\psi(X), \beta_\psi)$.*

Proof. The proof of Theorem 5.4.28 shows the necessity. If $f \mapsto \mathbb{E}^\mu[S_P f]$ is upper semi-continuous and does not attain ∞ on $(C_\psi(X), \beta_\psi)$ then by Fenchel-Moreau 5.A.9 it must be the infimum over a closed convex subset of the dual space. Mimicking the proof of Theorem 5.4.28 shows that all elements in this subset must satisfy the conditions of Theorem 5.4.23. $\square$

Definition 5.4.31. *For $f : X \to \overline{\mathbb{R}}$ and $\nu \in M^1_\psi(X)$ set*

$$T^\nu_P f = \inf_{g \in S_P^{-1} f} \mathbb{E}^\nu[g].$$

Lemma 5.4.32. *T^ν_P is sublinear.*

Proof. We have

$$\begin{aligned}\{h \in C_l(X) :\ S_P h \geq f+g\} &\supset \{h_1+h_2 :\ h_1,h_2 \in C_l(X),\ S_P h_1 + S_P h_2 \geq f+g\}\\ &\supset \{h_1+h_2 :\ h_1,h_2 \in C_l(X),\ S_P h_1 \geq f \ \text{ and } S_P h_2 \geq g\} =: B\end{aligned}$$

since S_P is superlinear. Hence

$$T^\nu_P(f+g) = \inf_{h \in S_P^{-1}(f+g)} \mathbb{E}^\nu[h] \leq \inf_{h \in B} \mathbb{E}^\nu[h] = T^\nu_P f + T^\nu_P g.$$

The positive homogeneity is obvious. $\square$

Theorem 5.4.33. *Let $P \subset Z$ be closed and satisfy Assumption 5.2.12 and let $\nu \in M^1_\psi(X)$. Assume ${}^\nu\mathcal{M}_\psi(P) \neq \emptyset$. Set $\mathcal{N}^\nu(P) := (p_1)_*({}^\nu\mathcal{M}_\psi(P))$ Then for any $f \in C_\psi(X)$ we have the following strong duality*

$$\sup_{\mu \in \mathcal{N}^\nu(P)} \mathbb{E}^\mu[f] = T^\nu_P f.$$

.

Proof. We show that $f \mapsto T^\nu_P f$ is lower semi-continuous. Let $f \in C_\psi(X)$ and $r \in \mathbb{R}$ be a pair such that $T^\nu_P(f) > r$ then we have to find an open set $\mathcal{U}$ such that

$$\inf_{g \in \mathcal{U}} T^\nu_P g > r.$$

Let $\mu \in {}^\nu\mathcal{M}_\psi(P)$ and let $\phi \in C_0(X)$ be from Lemma 5.4.27. Again set $\mathcal{U}_\epsilon := \{g \in C_\psi(X) : |f-g|\psi\phi < \epsilon\}$, then we have the upper bound

$$g(x) \geq f(x) - \frac{\epsilon}{\psi(x)\phi(x)}$$

for $g \in \mathcal{U}_\epsilon$. Hence by Theorem 5.4.15 for any $g \in \mathcal{U}_\epsilon$ we obtain

$$T^\nu_P g = \inf_{h \in S_P^{-1} g} \mathbb{E}^\nu[h] \geq \inf_{h \in S_P^{-1} g} \mathbb{E}^\mu[S_P h] \geq \mathbb{E}^\mu[g] \geq \mathbb{E}^\mu[f - \frac{\epsilon}{\psi\phi}] > -\infty$$

and thus for every $\epsilon > 0$ any $g \in \mathcal{U}_\epsilon$ has a lower integrable bound. By Lemma 5.4.17 for every convex function h we have $S_P h \geq h$ and thus in particular

$$S_P\left(\frac{\epsilon}{\phi\psi}\right) \geq \frac{\epsilon}{\phi\psi}.$$

It follows that for all $g \in \mathcal{U}_\epsilon$ we have for any $h \in S_P^{-1}(g)$

$$S_P\left(h + \frac{\epsilon}{\phi\psi}\right) \geq S_P h + S_P\left(\frac{\epsilon}{\phi\psi}\right) \geq f - \frac{\epsilon}{\phi\psi} + \frac{\epsilon}{\phi\psi} = f.$$

and hence $S_P^{-1}(g) + \frac{\epsilon}{\phi\psi} \subset S_P^{-1}(f)$. We obtain for all $g \in \mathcal{U}_\epsilon$

$$\begin{aligned} T_P^\nu g = \inf_{h \in S_P^{-1} g} \mathbb{E}^\nu[h] &= \inf_{h \in S_P^{-1} g} \mathbb{E}^\nu[h + \frac{\epsilon}{\phi\psi}] - \mathbb{E}^\nu[\frac{\epsilon}{\phi\psi}] \\ &\geq \inf_{h \in S_P^{-1} f} \mathbb{E}^\nu[h] - \mathbb{E}^\nu[\frac{\epsilon}{\phi\psi}] = T_P^\nu f - \mathbb{E}^\nu[\frac{\epsilon}{\phi\psi}] \end{aligned}$$

Thus there exists ϵ_0 such that

$$\inf_{g \in \mathcal{U}_{\epsilon_0}} T_P^\nu g > r.$$

Again by Theorem 5.A.9 we obtain a closed convex set $A \subset C_\psi(X)^* = M_\psi(X)$ such that

$$T_P^\nu f = \sup_{h^* \in A} h^*(f).$$

It remains to show that all elements in A are probability measures which satisfy the conditions of Theorem 5.4.23. We have

$$\sup_{h^* \in A} h^*(\mathbf{1}) = T_P^\nu \mathbf{1} = \inf_{g \in S_P^{-1}(\mathbf{1})} \mathbb{E}^\nu[g] \leq \mathbb{E}^\nu[\mathbf{1}] = 1$$

where the last inequality follows from $\mathbf{1} \in S_P^{-1}(\mathbf{1})$. Similarly we have

$$\sup_{h^* \in A} h^*(-\mathbf{1}) = T_P^\nu(-\mathbf{1}) = \inf_{g \in S_P^{-1}(-\mathbf{1})} \mathbb{E}^\nu[g] \leq \mathbb{E}^\nu[-\mathbf{1}] = -1$$

and hence we also have $\inf_{h^* \in A} h^*(\mathbf{1}) \geq 1$ implying that h has total mass 1. Assume there exists a measure in A which assigns negative mass to $B \subset X$, then there exists non negative $f \in C_\psi(X)$ with $h^*(f) < 0$ and we have

$$0 < \sup_{h^* \in A} h^*(-f) = T_P^\nu(-f) = \inf_{g \in S_P^{-1}(-f)} \mathbb{E}^\nu[g] \leq \mathbb{E}^\nu[0] = 0$$

since $0 \in S_P^{-1}(-f)$. This is a contradiction. For any $g \in C(X)$ which is is bounded below by k, we have $S_P g \geq k$ as well. For any $f \in C_b(X)$ with $f \leq S_P g$ we have $g \in S_P^{-1} f$

because of the previous inequality and thus for all $h^* \in A$ with $h^*(f) = \int f d\lambda$ we have

$$\mathbb{E}^\nu[g] \geq \mathbb{E}^\lambda[f].$$

Since this holds for all $f \in C_b(X)$ with $f \leq S_P g$ we obtain

$$\mathbb{E}^\nu[g] \geq \sup_{\substack{f \in C_b(X) \\ f \leq S_P g}} \mathbb{E}^\lambda[f].$$

Since by Assumption 5.2.12 $S_P g$ is lower semi-continuous and bounded below by k there exists an increasing sequence $(f_n)_n$ in $C_b(X)$ which converges pointwise to $S_P g$. Hence

$$\mathbb{E}^\lambda[S_P g] \geq \sup_{\substack{f \in C_b(X) \\ f \leq S_P g}} \mathbb{E}^\lambda[f] \geq \lim_n \mathbb{E}^\lambda[f_n] = \mathbb{E}^\lambda[S_P g]$$

and thus we have

$$\mathbb{E}^\nu[g] \geq \mathbb{E}^\lambda[S_P g].$$

Theorem 5.4.23 implies $A = \mathcal{N}^\nu(P)$. □

Corollary 5.4.34. *Under the assumptions of Theorem 5.4.33 the operator $f \mapsto T_P^\nu f$ is continuous with respect to $(C_\psi(X), \beta_\psi)$.*

Proof. By Theorem 5.4.33 we have $\sup_{\mathbb{Q} \in {}^\nu\mathcal{M}_\psi(P)} \mathbb{E}^\mathbb{Q}[f \circ p_1] = T_P^\nu f$. The result follows from Proposition 7.32 in [15] and the fact that ${}^\nu\mathcal{M}_\psi(P)$ is weak-$*$ compact and the map

$$(\mathbb{Q}, f) \mapsto \mathbb{E}^\mathbb{Q}[f \circ p_1]$$

is jointly continuous. □

5.4.4. Marginals of marginals

In this subsection, we explore the existence of martingale measures defined on a specified path space, where only the one-dimensional marginal distributions of the initial and final distributions are fixed. The motivation for this investigation, particularly from a financial perspective, is that when dealing with multiple assets, call option data provides the risk-neutral distribution of each individual asset price at specific times, but does not provide the joint risk-neutral distribution of the asset prices at a specific time step.
For completeness we state Theorem 7 in [71] in our notation.

Theorem 5.4.35 (Strassen)**.** *Let $X_1 \times \cdots \times X_n$ be Polish spaces and ψ_i be continuous positive and bounded functions on X_i. Set $\widetilde{X} = X_1 \times \cdots \times X_n$ and*

$$\frac{1}{\psi_{\widetilde{X}}} := \sum_i \frac{1}{\psi_i} \circ p_{X_i}$$

p_{X_i} denotes the projection onto X_i. Let Λ be a closed convex subset of probability measures $M_{\psi_{\widetilde{X}}}(\widetilde{X})$. Then there exists an element $\mathbb{Q} \in \Lambda$ with marginals μ_i for all $f_i \in C_{\psi_i}(X_i)$ bounded below we have

$$\sum_i \mathbb{E}^{\mu_i}[f_i] \geq \inf_{\lambda \in \Lambda} \sum_i \mathbb{E}^{\lambda}[f_i \circ p_{X_i}].$$

Proof. See Theorem 7 in [71] □

Definition 5.4.36. *For $X = \mathbb{R}^d$ denote by $x^i : X \to \mathbb{R}$ the projection onto the i-th component of X. In the following let $\gamma : \mathbb{R} \to \mathbb{R}$ be given by $\gamma(x) = \frac{1}{1+|x|^t}$ and set*

$$\frac{1}{\psi_X} = \sum_i \frac{1}{\gamma \circ x^i} = \frac{1}{d + \sum |x^i(x)|^t}.$$

Remark 5.4.37. *Notice that for any function $f \in C(X)$, $|f\psi_X|$ is bounded if and only if $|f\psi|$ is, since there exist constants $c_1, c_2 \in \mathbb{R}_{>0}$ such that $c_1\psi \leq \psi_X \leq c_2\psi$. Hence $C_{\psi_X}(X) = C_\psi(X)$ and $M_{\psi_X}(X) = M_\psi(X)$.*

Theorem 5.4.38. *Let $X = \mathbb{R}^d$. Given a weak-$*$-closed convex non empty set Λ of martingale probability measures in $M_{\psi_Z}(Z)$ and probability measures $\mu_i, \nu_i \in M^1_\gamma(\mathbb{R})$, $1 \leq i \leq d$ on $\mathbb{R}$, there exists a martingale measure in Λ such that the one dimensional marginal distributions of the initial marginal and final marginal are given by $(\mu_i)_i$ and $(\nu_i)_i$ respectively if and only if for all $f_i, g_i \in C_\psi(\mathbb{R})$, $1 \leq i \leq d$, we have*

$$\sum_{i=1}^d \mathbb{E}^{\nu_i}[f_i] + \mathbb{E}^{\mu_i}[g_i] \geq \inf_{\lambda \in \Lambda} \sum_{i=1}^d \mathbb{E}^{\lambda}[(\oplus_{i=1}^d f_i) \circ p_n + (\oplus_{i=1}^d g_i) \circ p_1] \tag{5.4.8}$$

where $\oplus_{i=1}^d f_i : X \to \mathbb{R}$ and $\oplus_{i=1}^d g_i : X \to \mathbb{R}$ are the direct sums of the $(f_i)_i$ and $(g_i)_i$ respectively.

Proof. Define $\psi_2 : X \times X \to \mathbb{R}$ by $\frac{1}{\psi_2(x,y)} := \frac{1}{\psi(x)} + \frac{1}{\psi(y)}$. We show first that the map $(p_1, p_n)_* : M_{\psi_Z}(Z) \to M_{\psi_2}(X \times X)$ mapping $\mathbb{Q} \mapsto (p_1, p_n)_*\mathbb{Q}$ the pushforward under $(p_1, p_n)_*$ maps Λ to a closed convex set of probability measures in $M^1_{\psi_2}(X \times X)$. The weak-$*$- topologies on $M^1_{\psi_2}(X \times X)$ and $M^1_{\psi_Z}(Z)$ are metrizable by the same argument as in Theorem 5.4.21. If $(\eta_i)_i$ is a sequence in $(p_1, p_n)_*(\Lambda) \subset M^1_{\psi_2}(X \times X)$ converging to η then by continuity of $q_2 : (X \times X \ni (x, y)) \mapsto y$ we have

$$\lim_i (q_2)_*(\eta_i) = (q_2)_*(\eta).$$

Since for each η_i there is $\mathbb{Q}_i \in \Lambda$ with

$$(p_1, p_n)_*(\mathbb{Q}_i) = \eta_i$$

we obtain

$$(p_n)_*(\mathbb{Q}_i) = (q_2)_*(\eta_i),$$

i.e. the second marginal of η_i is the final marginal of $\mathbb{Q}_i$. The convergence of $((q_2)_*(\eta_i))_i$ implies the weak-$*$ compactness of $((q_2)_*(\eta_i))_i$ in $M_\psi(X)$, i.e. the final marginal distribution $(\mathbb{Q}_i)_i$ are weak-$*$ compact. Mimicking the proof of Theorem 5.4.21 implies the convergence of a subsequence of $(\mathbb{Q}_i)_i$ to some $\mathbb{Q} \in \Lambda$ since Λ is closed. By continuity of $(p_1, p_n)_*$ we obtain $(p_1, p_n)_*(\mathbb{Q}) = \eta$ and thus $(p_1, p_n)_*(\Lambda)$ is closed. By Remark 5.4.37 applying Theorem 5.4.35 to the closed convex set $(p_1, p_n)_*\Lambda$ with $X_i = \mathbb{R}$, $\widetilde{X} = X \times X = \mathbb{R}^{2d}$ and $\psi_i = \gamma$ yields the result. $\square$

Corollary 5.4.39. *The inequality condition* (5.4.8) *of Theorem 5.4.38 holds if and only if*

$$\sum_i \mathbb{E}^{\nu_i}[f_i] + \mathbb{E}^{\mu_i}[g_i] \geq \inf_{\lambda \in \Lambda} \sum_i \mathbb{E}^{\lambda}[(\oplus_{i=1}^d f_i) \circ p_n + (\oplus_{i=1}^d g_i) \circ p_1].$$

holds for all $f_i, g_i \in C_\gamma(\mathbb{R}) \cap C_l(\mathbb{R})$.

Proof. For any $f_i, g_i \in C_\gamma(\mathbb{R})$, $f_i^k := \max(f_i, -k) \in C_\gamma(\mathbb{R}) \cap C_l(\mathbb{R})$ and $g_i^k := \max(g_i, -k) \in C_\gamma(\mathbb{R}) \cap C_l(\mathbb{R})$ converge pointwise to f_i and g_i respectively. Hence

$$\begin{aligned}\sum_i \mathbb{E}^{\nu_i}[f_i] + \mathbb{E}^{\mu_i}[g_i] &= \inf_k \sum_i \mathbb{E}^{\nu_i}[f_i^k] + \mathbb{E}^{\mu_i}[g_i^k] \\ &\geq \inf_k \inf_{\lambda \in \Lambda} \sum_i \mathbb{E}^{\lambda}[(\oplus_{i=1}^d f_i^k) \circ p_n + (\oplus_{i=1}^d g_i^k) \circ p_1] \\ &= \inf_{\lambda \in \Lambda} \sum_i \mathbb{E}^{\lambda}[(\oplus_{i=1}^d f_i) \circ p_n + (\oplus_{i=1}^d g_i) \circ p_1].\end{aligned}$$

$\square$

Theorem 5.4.40. *Let $P \subset X^n$ be closed and satisfy Assumption 5.2.12 and let μ_i, ν_i, $1 \leq i \leq d$ be probability distributions in $M_\gamma^1(\mathbb{R})$. Then there exists a martingale measure $\mathbb{Q} \in \mathcal{M}_\psi(P)$ with initial and final marginal μ and ν in $M_\psi^1(X)$ such that $\mu \in M_\psi^1(X)$ has marginals $\mu_i \in M_\gamma^1(\mathbb{R})$ and $\nu \in M_\psi^1(X)$ has marginals $\nu_i \in M_\gamma^1(\mathbb{R})$ for $1 \leq i \leq d$, if and only if for all $f_i \in C_\gamma(\mathbb{R}) \cap C_l(\mathbb{R})$ we have*

$$\sup_{\substack{\oplus\phi_i \leq S_P(\oplus f_i) \\ \phi_i \in C_b(\mathbb{R})}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] \leq \sum_i \mathbb{E}^{\nu_i}[f_i]. \tag{5.4.9}$$

Proof. We need to show that (5.4.9) together with Corollary 5.4.39 imply (5.4.8) of Theorem 5.4.38 with $\Lambda = \mathcal{M}_\psi(P)$. To this end assume that (5.4.9) is satisfied for all $f_i \in C_\gamma(\mathbb{R}) \cap C_l(\mathbb{R})$. Assume by contradiction that the set $\mathcal{M}_\psi(P)$ of martingale measures is empty, then, by the same argument as in Theorem 5.4.23 $S_P f \equiv \infty$ for any $f : X \to \mathbb{R}$ and hence

$$\infty > \sum_i \mathbb{E}^{\nu_i}[f_i] \geq \sup_{\substack{\oplus\phi_i \leq S_P(\sum \oplus f_i) \\ \phi_i \in C_b(\mathbb{R})}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] = \infty$$

for all $f_i \in C_\gamma(\mathbb{R}) \cap C_l(\mathbb{R})$. Set $f := \oplus f_i$ and $g := \oplus g_i$ for $f_i, g_i \in C_\gamma(\mathbb{R}) \cap C_l(\mathbb{R})$. We have

$$\sum_i \mathbb{E}^{\mu_i}[g_i] + \mathbb{E}^{\nu_i}[f_i] \geq \sum_i \mathbb{E}^{\mu_i}[g_i] + \sup_{\substack{\oplus\phi_i \leq S_P f \\ \phi_i \in C_b(\mathbb{R})}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] = \inf_{\mu \in \mathcal{C}(\mu_1,\dots,\mu_d)} \mathbb{E}^{\mu}[g + S_P f] \tag{5.4.10}$$

where $\mathcal{C}(\mu_1, \dots, \mu_d)$ denotes the set of couplings of $\mu_1, \dots, \mu_d$ and the last equality follows from the Kantorovich duality [72] since $S_P f$ is lower semi-continuous by Assumption 5.2.12 and bounded below by the assumptions on the f_i. By Proposition 5.4.12 we have

$$\begin{aligned} \inf_{\mu \in \mathcal{C}(\mu_1,\dots,\mu_d)} \mathbb{E}^{\mu}[g + S_P f] &\geq \inf_{x \in p_1(P)} g(x) + S_P f(x) \\ &= \inf_{x \in p_1(P)} [g(x) + \inf_{\nu \in \mathcal{D}_P^{\delta_x}(f \circ p_n)} \mathbb{E}^{\nu}[f \circ p_n]] \\ &\geq \inf_{\lambda \in \mathcal{M}_\psi(P)} \mathbb{E}^{\lambda}[f \circ p_n + g \circ p_1]. \end{aligned} \tag{5.4.11}$$

where

$$\inf_{\mu \in \mathcal{C}(\mu_1,\dots,\mu_d)} \mathbb{E}^{\mu}[g + S_P f] \geq \inf_{x \in p_1(P)} g(x) + S_P f(x)$$

holds since $S_P f \equiv \infty$ on $X \setminus p_1(P)$. Combining (5.4.10) and (5.4.11) we obtain

$$\sum_i \mathbb{E}^{\mu_i}[g_i] + \mathbb{E}^{\nu_i}[f_i] \geq \inf_{\lambda \in \mathcal{M}_\psi(P)} \mathbb{E}^{\lambda}[f \circ p_n + g \circ p_1]$$

which is condition (5.4.8) of Theorem 5.4.38 applied to $\Lambda = \mathcal{M}_\psi(P)$. Conversely we have for any martingale measure λ satisfying the constraints, by Theorem 5.4.15

$$\mathbb{E}^{\lambda_n}[f] \geq \mathbb{E}^{\lambda_1}[S_P f] \geq \sup_{\substack{\oplus\phi_i \leq S_P f \\ \phi_i \in C_b(\mathbb{R})}} \sum_i \mathbb{E}^{\mu_i}[\phi_i]$$

where λ_1 and λ_n in $M_\psi(X)$ denote the first and last marginal distribution with λ_1 having one-dimensional marginals given by $\mu_1, \dots, \mu_d$. $\square$

Proposition 5.4.41. *Assume $P \subset Z$ closed and satisfies Assumption 5.4.18. Denote by $\mathcal{M}_\psi(\mu_1, \dots, \mu_d, P) \subset \mathcal{M}_\psi(P)$ all martingale measures $\mathbb{Q} \subset \mathcal{M}_\psi(P)$ with $\mu := (p_1)_*(\mathbb{Q}) \in M^1_\psi(X)$ having fixed marginals $\mu_1, \dots, \mu_d \in M^1_\gamma(\mathbb{R})$. Set $\mathcal{M}^n_\psi(\mu_1, \dots, \mu_d, P)$ to be the set of final marginals of $\mathbb{Q} \in \mathcal{M}_\psi(\mu_1, \dots, \mu_d, P)$. If $\mathcal{M}_\psi(\mu_1, \dots, \mu_d, P) \neq \emptyset$, then we have for all $f_i \in C_\gamma(\mathbb{R})$*

$$\sup_{\substack{\oplus\phi_i \leq S_P(\oplus f_i) \\ \phi_i \in L^1(\mu_i)}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] = \inf_{\nu \in \mathcal{M}^n(\mu_1,\dots,\mu_d,P)} \sum_i \mathbb{E}^{\nu_i}[f_i].$$

Proof. We show that

$$F(\bigoplus f_i) := \sup_{\substack{\oplus\phi_i \le S_P(\oplus f_i)\\ \phi_i \in L^1(\mu_i)}} \sum_i \mathbb{E}^{\mu_i}[\phi_i]$$

is upper semi-continuous on $\bigoplus_{i=1}^d C_\gamma(\mathbb{R})$ equipped with the product γ-strict topology. Choose $\mathbb{Q} \in \mathcal{M}_\psi(\mu_1, \ldots, \mu_d, P)$ with initial distribution μ and final distribution ν. By Lemma 5.4.27 there exist $(\phi_i)_i$ such that $\frac{1}{\gamma\phi_i}$ has finite expectation under ν_i. For fixed $f_i \in C_\gamma(\mathbb{R})$ set $f := \bigoplus f_i \in C_\psi(X)$ and set $U_\epsilon(f) := \{\bigoplus g_i | \; \forall i : |f_i - g_i|\gamma\phi_i < \epsilon\}$. We have $\bigoplus g_i \le \bigoplus f_i + \frac{\epsilon}{\gamma\phi_i}$ for $\bigoplus g_i \in U_\epsilon(f)$ and by Theorem 5.4.15 $S_P(\bigoplus f + \frac{\epsilon}{\gamma\phi_i})$ is μ-integrable. Assume

$$\sup_{\substack{\sum \phi_i \le S_P(\bigoplus f_i)\\ \phi_i \in L^1(\mu_i)}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] < r$$

for some $r \in \mathbb{R}$. By Theorem 2.6 in [50] we have for $g \in U_\epsilon(f)$

$$\sup_{\substack{\sum \phi_i \le S_P(\bigoplus g_i)\\ \phi_i \in L^1(\mu_i)}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] = \inf_{\mathcal{C}(\mu_1,\ldots,\mu_d)} \mathbb{E}^{\mu}[S_P(\bigoplus g_i)] \le \inf_{\mathcal{C}(\mu_1,\ldots,\mu_d)} \mathbb{E}^{\mu}[S_P(\bigoplus f_i + \frac{\epsilon}{\gamma\phi_i})] < \infty$$

Since S_P is continuous for monotone decreasing sequences (see Lemma 5.4.7 together with Definition 5.4.9) and by monotone convergence we have

$$\inf_{\mathcal{C}(\mu_1,\ldots,\mu_d)} \mathbb{E}^{\mu}[S_P(\bigoplus f_i + \frac{\epsilon}{\gamma\phi_i})] \to \inf_{\mathcal{C}(\mu_1,\ldots,\mu_d)} \mathbb{E}^{\mu}[S_P(\bigoplus f_i)] = \sup_{\substack{\sum \phi_i \le S_P(\bigoplus f_i)\\ \phi_i \in L^1(\mu_i)}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] < r$$

as $\epsilon \to 0$. Hence there exists ϵ_0 such that

$$\sup_{\substack{\sum \phi_i \le S_P(\bigoplus g_i)\\ \phi_i \in L^1(\mu_i)}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] < r$$

for $g \in U_\epsilon(f)$ for all $0 < \epsilon \le \epsilon_0$. This shows that F is upper semi-continuous. F is superlinear since S_P is superlinear, and proper (see Definition 5.A) since $\mathcal{M}_\psi(\mu_1, \ldots, \mu_d, P)$ is non empty. Hence we have

$$\sup_{\substack{\oplus\phi_i \le S_P(\bigoplus f_i)\\ \phi_i \in L^1(\mu_i)}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] = \inf_{h^* \in A} h^*(f)$$

where A is a weakly closed convex set in the dual space $(\bigoplus_{i=1}^d C_\gamma(\mathbb{R}))^* = \bigoplus M_\gamma(\mathbb{R})$. Every functional in A is of the form

$$h^*(\oplus f_i) = \sum_i^d \int f_i d\nu_i$$

with $\nu_i \in M_\gamma(\mathbb{R})$. Using the property that $S_p f \ge c$ if $f \ge c$ and testing with $f =$

$0 \oplus \cdots \oplus \underbrace{f}_{i} \oplus \cdots \oplus 0$ together with the same arguments as in Theorem 5.4.28 yields that all ν_i are probability measures. Applying Theorem 5.4.40 we obtain $A = \{\bigoplus \nu_i | \exists \nu \in \mathcal{M}^n_\psi(\mu_1, \dots, \mu_d, P)$ s.t. $\forall i : \nu_i = (x^i)_* \nu\}$ where $x^i : (x_1 \dots, x_d) \in X \mapsto x_i$.

□

5.5. Proofs of the main theorems

Proof of Theorem 5.2.7. Choose $\psi(x) = \frac{1}{1+||x||}$, i.e. $t = 1$ in the definition of ψ. By Theorem 5.B.9 for any $f \in C(X)$ and closed $P \subset Z$ the map $S_P f$ is Borel measurable. We show that the inequality condition

$$\mathbb{E}^\mu[S_P f] \leq \mathbb{E}^\nu[f] \tag{5.5.1}$$

for $f \in C(X)$ with $f \geq 0$ implies that $S_P f$ is μ-integrable for $f \in C_\psi(X)$ and that the inequality condition (5.4.5) in Theorem 5.4.23 is satisfied. For any $f \in C_\psi(X)$ such that $f \geq -M$ for some $M \in \mathbb{R}$ we have $f + M \geq 0$ and hence by assumption

$$\mathbb{E}^\nu[f + M] \geq \mathbb{E}^\mu[S_P(f + M)] \geq \mathbb{E}^\mu[S_P f] + \mathbb{E}^\mu[S_P(M)] \geq \mathbb{E}^\mu[S_P f] + M$$

and thus $\mathbb{E}^\nu[f] \geq \mathbb{E}^\mu[S_P f]$ for all $f \in C_\psi(X)$ which are bounded from below. For $f \in C_\psi(X)$, the function $f\mathbb{1}_{\{f>0\}}$ is bounded below and by Lemma 5.4.17 we have $S_P(f\mathbb{1}_{\{f>0\}}) \geq 0$. By monotonicity of S_P we have $S_P f \leq S_P(f\mathbb{1}_{\{f>0\}})$ and thus

$$\mathbb{E}^\mu[(S_P f)\mathbb{1}_{\{S_P f>0\}}] \leq \mathbb{E}^\mu[S_P(f\mathbb{1}_{\{f>0\}})] \leq \mathbb{E}^\nu[f\mathbb{1}_{\{f>0\}}] < \infty$$

where the second to last inequality follows from (5.5.1). This shows that $S_P f$ is μ-integrable. Any $f \in C_\psi(X)$ can be approximated by the sequence $f_n = \max(f, -n)$ and hence by Lemma 5.4.8 we have

$$\infty > \mathbb{E}^\nu[f] = \lim_n \mathbb{E}^\nu[f_n] \geq \lim_n \mathbb{E}^\mu[S_P f_n] = \mathbb{E}^\mu[S_P f].$$

Thus the inequality condition (5.4.5) of Theorem 5.4.23 is satisfied which yields the existence of the desired martingale measure.
Conversely if there exists a martingale measure $\mathbb{Q}$ supported on $P \subset Z$, then by Theorem 5.4.15 we have for all $f \in C(X)$ which are ν-integrable that

$$\mathbb{E}^\mu[S_P f] \leq \mathbb{E}^\nu[f].$$

If $f \in C(X)$ with $f \geq 0$ and f not ν-integrable then $\mathbb{E}^\nu[f] = \infty$ and the inequality $\mathbb{E}^\nu[f] \geq \mathbb{E}^\mu[S_P f]$ trivially holds. □

Proof of Theorem 5.2.9. For any $f \in \mathcal{F}_\psi(X)$ there exists a decreasing sequence $(f_i)_i$ in

$C_\psi(X)$ such that $\lim_i f_i = f$ pointwise. Applying Theorem 5.4.28 to each f_i we obtain

$$\begin{aligned}\inf_{\mathbb{Q}\in{}_\mu\mathcal{M}_\psi(P)} \mathbb{E}^{\mathbb{Q}}[f\circ p_n] &= \inf_{\mathbb{Q}\in{}_\mu\mathcal{M}_\psi(P)} \inf_i \mathbb{E}^{\mathbb{Q}}[f_i\circ p_n] \\ &= \inf_i \inf_{\mathbb{Q}\in{}_\mu\mathcal{M}_\psi(P)} \mathbb{E}^{\mathbb{Q}}[f_i\circ p_n] \\ &= \inf_i \mathbb{E}^\mu[S_P f_i] = \mathbb{E}^\mu[S_P f]\end{aligned}$$

where the last equality follows from Lemma 5.4.8. □

Proof of Theorem 5.2.14. For any $f \in \mathcal{F}_\psi(X)$ and any $h \in S_P^{-1}f$ we have by Theorem 5.4.15 that for all $\mathbb{Q} \in {}^\nu\mathcal{M}_\psi(P)$

$$\mathbb{E}^\nu[h] \geq \mathbb{E}^{\mathbb{Q}}[S_P h \circ p_1] \geq \mathbb{E}^{\mathbb{Q}}[f\circ p_1].$$

Taking the infimum on the left hand side with respect to $h \in S_P^{-1}f$ and the supremum on the right hand side with respect to $\mathbb{Q} \in {}^\nu\mathcal{M}_\psi(P)$ we obtain

$$T_P^\nu f \geq \sup_{\mathbb{Q}\in{}^\nu\mathcal{M}_\psi(P)} \mathbb{E}^{\mathbb{Q}}[f\circ p_1]$$

Again for any $f \in \mathcal{F}_\psi(X)$ there exists a decreasing sequence $(f_i)_i$ in $C_\psi(X)$ such that $\lim_i f_i = f$ pointwise. Denote $Y := \{g \in C_\psi(X) : \ g \geq f\}$, then Y is convex. It follows that $f(x) = \inf_{g\in Y} g(x)$. By Sion's minimax theorem, see [68] or [5], applied to the bilinear form $\mathbb{E}^{(\cdot)}[(\cdot)\circ p_1]$ we obtain

$$\sup_{\mathbb{Q}\in{}^\nu\mathcal{M}_\psi(P)} \inf_{g\in Y} \mathbb{E}^{\mathbb{Q}}[g\circ p_1] = \inf_{g\in Y} \sup_{\mathbb{Q}\in{}^\nu\mathcal{M}_\psi(P)} \mathbb{E}^{\mathbb{Q}}[g\circ p_1].$$

since by Corollary 5.4.22 ${}^\nu\mathcal{M}_\psi(P)$ is compact. Hence applying Theorem 5.4.33 to each g yields

$$\begin{aligned}\sup_{\mathbb{Q}\in{}^\nu\mathcal{M}_\psi(P)} \mathbb{E}^{\mathbb{Q}}[f\circ p_1] &= \sup_{\mathbb{Q}\in{}^\nu\mathcal{M}_\psi(P)} \inf_{g\in Y} \mathbb{E}^{\mathbb{Q}}[g\circ p_1] \\ &= \inf_{g\in Y} \sup_{\mathbb{Q}\in{}^\nu\mathcal{M}_\psi(P)} \mathbb{E}^{\mathbb{Q}}[g\circ p_1] \\ &= \inf_{g\in Y} T_P^\nu(g) \geq T_P^\nu(f).\end{aligned}$$

Combining the two inequalities yields the result. □

Proof of Theorem 5.2.16. We show that the inequality condition implies the existence condition (5.4.9) of Theorem 5.4.40. To that end, we have for any $(f_i)_i$ with $f_i \in$

$C_\gamma(\mathbb{R}) \cap C_l(\mathbb{R})$ there exists M such that $f_i + M \geq 0$ for all i and hence

$$\begin{aligned}
\sup_{\substack{\oplus\phi_i \leq S_P(\oplus f_i)= \\ \phi_i \in C_b(\mathbb{R})}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] &= \sup_{\substack{\oplus\phi_i \leq S_P(\oplus f_i) \\ \phi_i \in C_b(\mathbb{R})}} \sum_i \mathbb{E}^{\mu_i}[\phi_i + M] - Md \\
&= \sup_{\substack{\oplus\phi_i \leq S_P(\oplus f_i)+Md \\ \phi_i \in C_b(\mathbb{R})}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] - Md \\
&\leq \sup_{\substack{\oplus\phi_i \leq S_P(\oplus f_i+Md) \\ \phi_i \in C_b(\mathbb{R})}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] - Md \\
&\leq \sum_i \mathbb{E}^{\nu_i}[f_i + M] - Md = \sum_i \mathbb{E}^{\nu_i}[f_i]
\end{aligned}$$

and thus the inequality condition of Theorem 5.4.40 is satisfied. Conversely assume that there exists a $\mathbb{Q} \in \mathcal{M}_\psi(P)$ such that $\mathbb{Q}$ has initial and final marginal μ, ν which have marginals $(\mu_i)_i$ and $(\nu_i)_i$. For $f_i \in C(\mathbb{R})$ with $f_i \geq 0$ we have by Theorem 5.4.15

$$\sup_{\substack{\oplus\phi_i \leq S_P(\oplus f_i) \\ \phi_i \in C_b(\mathbb{R})}} \sum_i \mathbb{E}^{\mu_i}[\phi_i] \leq \mathbb{E}^{\mu}[S_P(\oplus f_i)] \leq \mathbb{E}^{\nu}[\oplus f_i] = \sum_i \mathbb{E}^{\nu_i}[f_i].$$

The duality result is just Proposition 5.4.41. □

5.6. Proof of absence of arbitrage condition

Set $V = \mathbb{R}^d_{>0} \times \mathbb{R} \times \mathbb{R}^d_{>0}$.

Definition 5.6.1. *For $(s_1, c, s_2) \in V$ define the projections*

$$p_1(s_1, c, s_2) := s_1, \quad p_C(s_1, c, s_2) := c, \quad p_2 := (s_1, c, s_2) = s_2.$$

Theorem 5.6.2. *Let $L : \mathbb{R}^d_{>0} \times \mathbb{R}^d_{>0} \to \mathbb{R}$ be continuous and $\psi_1 : \mathbb{R}^d_{>0} \to \mathbb{R}$ satisfy Assumption 5.3.5 and set $\psi_2 : \mathbb{R} \to \mathbb{R},\ x \mapsto \frac{1}{1+|x|}$. Denote $\mathcal{P}_L := \cup \mathcal{P}_L(\mu_1, \mu_C, \mu_2)$ where the union is taken over all $(\mu_1, \mu_C, \mu_2) \in M^1_{\psi_1}(\mathbb{R}^d_{>0}) \times M^1_{\psi_2}(\mathbb{R}) \times M^1_{\psi_1}(\mathbb{R}^d_{>0})$. Define $\psi_L : V \to \mathbb{R}$ via*

$$\frac{1}{\psi_L} = \frac{1}{\psi_1 \circ p_1} + \frac{1}{\psi_1 \circ p_2} + \frac{1}{\psi_2 \circ p_C}.$$

Then $\mathcal{P}_L$ is weak-$*$-closed and convex with respect to the topology generated by $C_{\psi_L}(V)$. Furthermore $\mathcal{P}_L(\mu_1, \mu_C, \mu_2) \neq \emptyset$ if and only if for all $f, g \in C_{\psi_1}(\mathbb{R}^d_{>0})$ and $h \in C_{\psi_2}(\mathbb{R})$ which are bounded from below we have

$$\mathbb{E}^{\mu_1}[f] + \mathbb{E}^{\mu_2}[g] + \mathbb{E}^{\mu_C}[h] \geq \inf_{\lambda \in \mathcal{P}_L} \mathbb{E}^{\lambda}[f \circ p_1 + g \circ p_2 + h \circ p_C]. \tag{5.6.1}$$

Proof. Note that $\mu \in \mathcal{P}_L$ if and only if for all bounded $f \in C_b(\mathbb{R}^d_{>0} \times \mathbb{R})$ we have

$$\mathbb{E}^\mu[(p_2 - p_1) f \circ (p_1, p_C)] = 0, \quad \mathbb{E}^\mu[(L \circ (p_1, p_2) - p_C) f \circ (p_1, p_C)] = 0.$$

Since all functions under the expectation are in $C_\psi(V)$, $\mathcal{P}_L$ is clearly closed. The second statement is just an application of Theorem 5.4.35 on the Polish spaces $X_1 = \mathbb{R}^d_{>0}$, $X_2 = \mathbb{R}$, $X_3 = \mathbb{R}^d_{>0}$ with $\Lambda = \mathcal{P}_L$. $\square$

Proof of Theorem 5.3.7. Set $\psi_1 = \psi$ and $\psi_2(x) = \frac{1}{1+|x|}$. We first show that (5.3.1) implies (5.6.1) in Theorem 5.6.2. Let $f, g \in C_{\psi_1}(\mathbb{R}^d_{>0}) \cap C_l(\mathbb{R}^d_{>0})$ and $h \in C_{\psi_2}(\mathbb{R}) \cap C_l(\mathbb{R})$ and assume w.l.o.g. that $f \geq 0$, as otherwise one can just add constants.

$$\begin{aligned}
\mathbb{E}^{\mu_2}[f] + \mathbb{E}^{\mu_1}[g] + \mathbb{E}^{\mu_C}[h] &\geq \sup_{\substack{\varphi_1 + \varphi_C \leq S_P f \\ \varphi_1 \in C_b(\mathbb{R}^d_{>0}), \varphi_C \in C_b(\mathbb{R})}} \mathbb{E}^{\mu_1}[\varphi_1 + g] + \mathbb{E}^{\mu_C}[\varphi_C + h] \\
&= \sup_{\substack{\varphi_1 + \varphi_C \leq S_P f + g + h \\ \varphi_1 \in C_b(\mathbb{R}^d_{>0}), \varphi_C \in C_b(\mathbb{R})}} \mathbb{E}^{\mu_1}[\varphi_1] + \mathbb{E}^{\mu_C}[\varphi_C] \qquad (5.6.2) \\
&= \inf_{\eta \in \mathcal{C}(\mu_1, \mu_C)} \mathbb{E}^\eta[S_P f] + \mathbb{E}^\eta[g \circ p_1] + \mathbb{E}^\eta[h \circ p_c]
\end{aligned}$$

where the first inequality follows from (5.3.1) and the last equality from an application of Theorem 2.6 and Proposition 1.33 in [50] to $S_P f + g + h$ since it is bounded below and lower semi-continuous by Assumption 5.3.6. Set $W := (\mathbb{R}^d_{>0} \times \mathbb{R}) \times (\mathbb{R}^d_{>0} \times \mathbb{R})$ and

$$\frac{1}{\psi_W(s_1, c_1, s_2, c_2)} = \frac{1}{\psi_1(s_1)} + \frac{1}{\psi_2(c_1)} + \frac{1}{\psi_1(s_2)} + \frac{1}{\psi_2(c_2)}.$$

Notice that we have $P \subset W$. Under the pushforward of the map

$$G : W \to V, (s_1, c_1, s_2, c_2) \mapsto (s_1, c_1, s_2)$$

we have $G_*(\mathcal{M}_{\psi_W}(P)) \subset \mathcal{P}_L$. On W define $q_2 : (s_1, c_1, s_2, c_2) \mapsto (s_2, c_2)$. Then applying Proposition 5.4.12 we obtain

$$\begin{aligned}
\inf_{\eta \in \mathcal{C}(\mu_1, \mu_C)} \mathbb{E}^\eta[S_P f] + \mathbb{E}^\eta[g \circ p_1] + \mathbb{E}^\lambda[h \circ p_c] &\geq \inf_{(s_1, c) \in \mathbb{R}^d_{>0} \times \mathbb{R}} S_P f(s_1, h) + g(s_1) + h(c) \\
&\geq \inf_{(s_1, c) \in \mathbb{R}^d_{>0} \times \mathbb{R}} \inf_{\rho \in \mathcal{D}_P^{\delta_{(s_1, c)}}(\widehat{f} \circ q_2)} \mathbb{E}^\rho[\widehat{f} \circ q_2] + g(s_1) + h(c) \qquad (5.6.3) \\
&\geq \inf_{\lambda \in \mathcal{P}_L} \mathbb{E}^\lambda[f \circ p_2 + g \circ p_1 + h \circ p_C]
\end{aligned}$$

where the last inequality follows from $G_*(\mathcal{D}_P^{\delta_{(s_1,c)}}(\widehat{f} \circ q_2)) \subset G_*(\mathcal{M}_{\psi_W}(P)) \subset \mathcal{P}_L$. Combining (5.6.2) and (5.6.3) we obtain

$$\mathbb{E}^{\mu_2}[f] + \mathbb{E}^{\mu_1}[g] + \mathbb{E}^{\mu_C}[h] \geq \inf_{\lambda \in \mathcal{P}_L} \mathbb{E}^{\lambda}[f \circ p_2 + g \circ p_1 + h \circ p_C]$$

Applying Corollary 5.4.39 implies (5.6.1) in Theorem 5.6.2. Conversely every $\lambda \in \mathcal{P}_L(\mu_1, \mu_C, \mu_2)$ induces a martingale measure $\mathbb{Q} \in \mathcal{M}_{\psi_W}(P)$ on W. Thus for all $f \in C_{\psi_1}(\mathbb{R}^d_{>0}) \cap C_l(\mathbb{R}^d)$ we must have by Proposition 5.4.15

$$\mathbb{E}^{\mu_2}[f] = \mathbb{E}^{\lambda}[f \circ p_2] \geq \mathbb{E}^{\lambda}[S_P f \circ (p_1, p_C)] \geq \sup_{\substack{\varphi_1 + \varphi_C \leq S_P f \\ \varphi_1 \in C_b(\mathbb{R}^d_{>0}), \varphi_C \in C_b(\mathbb{R})}} \mathbb{E}^{\mu_1}[\varphi_1] + \mathbb{E}^{\mu_C}[\varphi_C].$$

□

For the rest of the section we set $d = 1$ and

$$L(s_1, s_2) := -\frac{2}{\tau} \log\left(\frac{s_2}{s_1}\right).$$

where $\tau > 0$ and

$$\frac{1}{\psi(x)} := 1 + |x| + |\log(x)|.$$

Notice that $L(s_1, s_1) = 0$.

Proposition 5.6.3. *For $f \in C(\mathbb{R}_{>0})$ we have*

$$S_P f(s_1, c) = \begin{cases} \sup\{as_1 + bc + d : \ as_2 + bL(s_1, s_2) + d \leq f(s_2) \text{ for all } s_2 \in \mathbb{R}_{>0}, \ a, b, d \in \mathbb{R}\} & c > 0 \\ f(s_1) & c = 0 \\ \infty & \text{else.} \end{cases}$$

Proof. Fix $s_1 \in \mathbb{R}_{>0}$. Set $G := \{(x, L(s_1, x)) : \ x \in \mathbb{R}_{>0}\}$, the graph of $x \mapsto L(s_1, x)$. We have

$$\begin{aligned} S_P f(s_1, c) &= \sup\{g(s_1, c) : \ g \text{ convex and } g(s_2, L(s_1, s_2)) \leq f(s_2) \text{ for all } s_2 \in \mathbb{R}_{>0}\} \\ &\geq \sup\{g(s_1, c) : \ g \text{ affine and } g(s_2, L(s_1, s_2)) \leq f(s_2) \text{ for all } s_2 \in \mathbb{R}_{>0}\} =: \tilde{f}(s_1, c) \end{aligned}$$

since any affine function is convex. If $c > L(s_1, s_1) = 0$ then the point (s_1, c) is in the interior of $\mathrm{conv}(G)$. By abuse of notation set $\overline{f}_G(s_1, c) := \overline{(\widehat{f})}_G(s_1, c)$ We have $\overline{f}_G(s_1, c) = S_P f(s_1, c)$ and $\overline{f}_G$ is convex. If $S_P f(s_1, c) = -\infty$ then $\tilde{f}(s_1, c) = -\infty$. For any $(x, y) \in \mathrm{conv}(G)^\circ$ there exist three points $(x_i, L(s_1, x_i))$, $i = 1, 2, 3$ and $\alpha_i > 0$ with $\sum \alpha_i = 1$ such that

$$(x, y) = \sum \alpha_i (x_i, L(s_1, x_i))$$

since in $\mathbb{R}^2$ every point inside the convex hull of a set is the convex combination of at most three points. Hence by Proposition 5.4.3 we have

$$\overline{f}_G(x,y) \le \sum \alpha_i \widehat{f}(x_i, L(s_1,x_i)) = \sum \alpha_i f(x_i) < \infty$$

since $\mu = \sum \alpha_i \delta_{(x_i, L(s_1,x_i))} \in \Xi_G^{(x,y)}(\widehat{f})$. Assume now that $S_P f(s_1,c) = \overline{f}_G(s_1,c) > -\infty$ then by the same argument as in Lemma 5.C.3 $\overline{f}_G$ is real valued on $\operatorname{conv}(G)^\circ$. Since $\overline{f}_G$ is convex it follows that it is continuous on $\operatorname{conv}(G)^\circ$. Thus by the supporting hyperplane theorem there exists an affine g such that $g(x,y) \le \overline{f}_G(x,y)$ and $g(s_1,c) = \overline{f}_G(s_1,c)$. Thus $S_P f(s_1,c) = \widetilde{f}(s_1,c)$. If $S_P f(s_1,c) = \overline{f}_G(s_1,c) = -\infty$ then by $S_P f(s_1,c) \ge \widetilde{f}(s_1,c)$ equality must hold. This proves the case $c > L(s_1,s_1)$. We proceed now to $c = L(s_1,s_1) = 0$. By Proposition 5.4.3 we have

$$\overline{f}_G(x,y) = \inf_{\lambda \in \Xi_G^{(x,y)}(\widehat{f})} \mathbb{E}^\lambda[\widehat{f}].$$

Since $L(s_1,\cdot)$ is strictly convex we have for $c = L(s_1,s_1)$ that $\Xi_G^{(s_1,c)}(f) = \{\delta_{(s_1,c)}\}$ and hence $\overline{f}_G(s_1,c) = S_P f(s_1,c) = \widehat{f}(s_1,c) = f(s_1)$. If $c < L(s_1,s_1)$ then $\Xi_G^{(s_1,c)}(f) = \emptyset$ and hence $S_P f(s_1,c) = \infty$. □

Proposition 5.6.4. *We have*

$$S_P f(s_1,c) = \sup\{as_1 + bc + d :\ as_2 + bL(s_1,s_2) + d \le f(s_2) \text{ for all } s_2 \in \mathbb{R}_{>0},\ a,b,d \in \mathbb{R}\}$$

for all $f \in C_\psi(\mathbb{R}_{>0})$.

Proof. By Proposition 5.6.3 we must show that for

$$\widetilde{f}(s_1,c) := \sup\{as_1 + bc + d :\ as_2 + bL(s_1,s_2) + d \le f(s_2) \text{ for all } s_2 \in \mathbb{R}_{>0},\ a,b,d \in \mathbb{R}\}$$

we have $\widetilde{f}(s_1,c) = f(s_1)$ for $c = L(s_1,s_1) = 0$ and $\widetilde{f}(s_1,c) = \infty$ for $c < L(s_1,s_1)$ in the special case $f \in C_\psi(\mathbb{R}_{>0})$. In the following fix $f \in C_\psi(\mathbb{R}_{>0})$ and $s_1 > 0$. Set $l(x) := -\frac{2}{\tau}\log(x/s_1) = L(s_1,x)$. The tangent of l at s_1 is given by $l(s_1) - t\frac{2}{\tau s_1}$. The strategy of the proof is to construct explicit affine functions g^ϵ which satisfy $f(x) \ge g^\epsilon(x, l(x))$ and such that $\lim_{\epsilon\to 0} g^\epsilon(s_1, l(s_1)) = f(s_1)$. This, then implies $\widetilde{f}(s_1, l(s_1)) \ge f(s_1)$. Together with $\widetilde{f}(s_1, l(s_1)) \le S_P f(s_1, l(s_1)) = f(s_1)$, this forces equality. Set

$$g_M^\epsilon(x,z) = f(s_1) - \epsilon - M(z - l(s_1) + \frac{2}{\tau s_1}(x - s_1)).$$

Clearly g_M^ϵ is affine and we have $g_M^\epsilon(s_1, L(s_1,s_1)) = f(s_1) - \epsilon$. Notice by strict convexity of l and the mean value theorem that $(l(x) - l(s_1) + \frac{2}{\tau s_1}(x - s_1)) \ge 0$ is always non-negative and strictly positive for $x \ne s_1$. Next we show that for any $0 < \epsilon$ there exists $M > 0$ such that

$$g_M^\epsilon(x, l(x)) \le f(x).$$

For $x < 1$ we have

$$g_M^\epsilon(x, l(x)) = -Ml(x) + f(s_1) - \epsilon - M(-l(s_1) + \frac{2}{\tau s_1}(x - s_1)) \leq -Ml(x) + C_1$$

for some appropriate constant which satisfies $C_1 > f(s_1) - \epsilon + M(l(s_1) + \frac{2}{\tau})$. Since $f \in C_\psi(\mathbb{R}^d)$ there is a constant $R > 0$ such that

$$|f(x)| < R(1 + |x| + |\log(x)|).$$

Moreover,

$$\frac{|Ml(x) - C_1|}{R(1 + |x| + |\log(x)|)} \to \frac{2M}{\tau R} \quad \text{as } x \to 0.$$

Hence choosing $M_1 > \frac{\tau R}{2}$ there exists $1 > k_1 > 0$ such that for $x \in (0, k_1)$ we have

$$g_{M_1}^\epsilon(x, l(x)) < -R(1 + |x| + |\log(x)|) < f(x).$$

Similarly for $x > 1$ we have

$$g_M^\epsilon(x, l(x)) = -M\frac{2}{\tau s_1}x + f(s_1) - \epsilon - M(l(x) - l(s_1) - \frac{2}{\tau}) \leq -M\frac{2}{\tau s_1}x + C_2 x$$

where $C_2 > \frac{1}{x}(f(s_1) - \epsilon - M(l(x) - l(s_1) - \frac{2}{\tau}))$ for all $x > 1$. Again we obtain

$$\frac{|2M/(\tau s_1)x - C_2 x|}{R(1 + |x| + |\log(x)|} \to \frac{|2M/(\tau s_1) - C_2|}{R} \quad \text{as } x \to \infty.$$

Hence choosing $M_2 > \tau s_1 \frac{R + C_2}{2}$ we obtain $k_2 > 1$ such that for all $x \in (k_2, \infty)$ we have

$$g_{M_2}^\epsilon(x, l(x)) < -R(1 + |x| + |\log(x)|) < f(x).$$

Since f is continuous at s_1 there exists as small neighbourhood $(s_1 - \delta, s_1 + \delta)$ such that for $x \in (s_1 - \delta, s_1 + \delta)$ we have $f(x) > f(s_1) - \epsilon$ and hence by non negativeness of $(l(x) - l(s_1) + \frac{2}{\tau s_1}(x - s_1))$ we have

$$f(x) > g_M^\epsilon(x, l(x))$$

for $x \in (s_1 - \delta, s_1 + \delta)$ and arbitrary $M \geq 0$. Finally consider the set $K := [k_1, k_2] \setminus (s_1 - \delta, s_1 + \delta)$. Clearly K is compact, since it is closed and bounded. Hence the continuous function

$$g_M^\epsilon(x, l(x)) - f(x) = f(s_1) - f(x) - \epsilon - M(l(x) - l(s_1) + \frac{2}{\tau s_1}(x - s_1)) < D$$

for some $D \in \mathbb{R}$ and $l(x) - l(s_1) + \frac{2}{\tau s_1}(x - s_1)$ must be bounded from below by some $h > 0$ on K. Thus

$$g_{M + \widetilde{M}}^\epsilon(x, l(x)) - f(x) \leq D - \widetilde{M}h.$$

Choose $\widetilde{M}$ large enough such that the right hand side is negative and set $M_3 = M + \widetilde{M}$. Choosing now $M_0 := \max(M_1, M_2, M_3)$ we obtain

$$f(x) \geq g^{\epsilon}_{M_0}(x, l(x))$$

and $g^{\epsilon}_{M_0}(s_1, l(s_1)) = g^{\epsilon}_{M_0}(s_1, L(s_1, s_1)) = f(s_1) - \epsilon$. Letting $\epsilon \to 0$ we obtain $\widetilde{f}(s_1, L(s_1, s_1)) \geq f(s_1)$ and since $\widetilde{f}(s_1, c) \leq S_P f(s_1, c)$ equality must hold by Proposition 5.6.3. This proves the case $c = L(s_1, s_1) = 0$. For $c < L(s_1, s_1)$ choose an arbitrary $\epsilon > 0$ and its accompanying M_0. For all $M > M_0$ we have $f(x) \geq g^{\epsilon}_M(x, l(x))$ and

$$g^{\epsilon}_M(s_1, c) = f(s_1) - \epsilon - M(c - l(s_1)).$$

Since $c - l(s_1) = c - L(s_1, s_1) = c < 0$ and letting $M \to \infty$ we obtain $\widetilde{f}(s_1, c) = \infty$. $\square$

Lemma 5.6.5. *For all $f \in C_\psi(\mathbb{R}_{>0})$ $S_P f$ is lower semi-continuous and we have the identity*

$$S_P f(s_1, c) = S_P f(1, c - \frac{2}{\tau} \log(s_1)).$$

Proof. Set $O^{s_1}_P(f) := \{g : \ g \text{ affine and } g(x, L(s_1, x)) \leq f(x) \text{ for } x > 0\}$. By Proposition 5.6.4 we have

$$S_P f(s_1, c) = \sup_{g \in O^{s_1}_P(f)} g(s_1, c).$$

Since every element in $O^{s_1}_P(f)$ is lower semi-continuous, so is its pointwise supremum and hence $S_P f$ is lower semi-continuous in its second argument. Notice that if $g \in O^1_P(f)$ then the affine map $g_{s_1}(x, c) = g(x, c - \frac{2}{\tau} \log(s_1))$ satisfies $g_{s_1} \in O^{s_1}_P(f)$. Conversely for $g \in O^{s_1}_P(f)$ the function $g^{s_1}(x, c) = g(x, c + \frac{2}{\tau} \log(s_1))$ satisfies $g^{s_1} \in O^1_P(f)$. The identity readily follows. From

$$S_P f(s_1, c) = S_P f(1, c - \frac{2}{\tau} \log(s_1))$$

it follows that S_P is jointly lower semi-continuous since log is continuous on $\mathbb{R}_{>0}$. $\square$

Lemma 5.6.6. *Let X be locally compact Polish space and let $\mu \in M_\psi(X)$, then there exists $\phi : X \mapsto (0, \infty)$ in $C_0(X)$ such that*

$$\mathbb{E}^\mu \left[\frac{1}{\psi \phi} \right] < \infty.$$

Proof. Since X is a Radon space μ is regular and thus there exists a sequence of compact sets $(U_n)_n$ with $U_n \subset U_{n+1}$ such that

$$\lim_{n \to \infty} \mu(U_n) = 1.$$

Let $K_1 \supset U_1$ be a compact neighbourhood of U_1 and inductively define K_n to be a neighbourhood of $U_n \cup K_{n-1}$. Since $K_n \supset U_n$ we have $\lim_{n \to \infty} \mu(K_n) = 1$. W.l.o.g. assume that $\mathbb{E}^\mu[\frac{1}{\psi} \mathbb{1}_{X \setminus K_n}] \leq \mathbb{E}^\mu[1/\psi]/4^n$ for otherwise we can pass onto a sub-sequence.

By Urysohn's lemma there exists a continuous function $f_n : X \to [0,1]$ such that $f_n \equiv 1$ on K_n and $f_n \equiv 0$ on $X \setminus K_{n+1}^\circ$. Set

$$\phi = \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} f_n.$$

Notice that $f_k \equiv 0$ on $X \setminus K_n$ for $1 \leq k \leq n-1$. Hence on $X \setminus K_n$ we have

$$\phi \leq \sum_{k=n}^{\infty} \frac{1}{2^{k+1}} f_k \leq \sum_{k=n}^{\infty} \frac{1}{2^{k+1}} = \frac{1}{2^n}.$$

This implies $\phi \in C_0(X)$ and $\phi \leq 1/2^n$ on $K_{n+1} \setminus K_n$. We have

$$\begin{aligned}
&\mathbb{E}^\mu[\frac{1}{\psi\phi}\mathbb{1}_{\cup K_n}] = \mathbb{E}^\mu[\frac{1}{\psi\phi}\sum \mathbb{1}_{K_{n+1}\setminus K_n}] = \sum \mathbb{E}^\mu[\frac{1}{\psi\phi}\mathbb{1}_{K_{n+1}\setminus K_n}] \\
&\leq \sum 2^n \sum \mathbb{E}^\mu[\frac{1}{\psi}\mathbb{1}_{K_{n+1}\setminus K_n}] \leq \sum 2^n \mathbb{E}^\mu[\frac{1}{\psi}\mathbb{1}_{X\setminus K_n}] \\
&\leq \sum \frac{1}{2^n}\mathbb{E}^\mu[\frac{1}{\psi}] = \mathbb{E}^\mu[\frac{1}{\psi}] < \infty.
\end{aligned}$$

□

Proof of Lemma 5.3.9. By Propositions 5.6.3 we have

$$S_P f(s_1, L(s_1, s_1)) = S_P f(s_1, 0) = f(s_1)$$

and $S_P f(s_1, c) = \infty$ for $c < 0$. Applying Proposition 5.6.4 and Lemma 5.6.5 yields the result. □

Proof of Theorem 5.3.10. Clearly the set $P = \{((s_1, c), (s_2, L(s_1, s_2))) : s_1, s_2 \in \mathbb{R}_{>0},\ c \in \mathbb{R}\} \subset \mathbb{R}_{>0} \times \mathbb{R} \times \mathbb{R}_{>0} \times \mathbb{R}$ is closed since $L(s_1, s_2) = -\frac{2}{\tau}\log(s_2/s_1)$ is continuous on $\mathbb{R}_{>0} \times \mathbb{R}_{>0}$. Assumption 5.3.5 is clearly satisfied for $1/\psi(x) = 1 + |x| + |\log(x)|$. Lemma 5.6.5 combined with Lemma 5.6.4 shows that P satisfies Assumption 5.3.6 and that we have

$$\widetilde{f}(c - \frac{2}{\tau}\log(s)) = S_P f(s, c).$$

Hence we can apply Theorem 5.3.7. This shows that *2.* and *3.* are equivalent. The equivalence of *2.* and *1.* is shown in Theorem 6 in [42]. The proof of the duality result is analogous to the proof of Proposition 5.4.41 by showing that the operator

$$F(f) := \sup_{\substack{\phi_1 + \phi_{\mathrm{VIX}^2} \leq S_P f \\ \phi_1 \in L^1(\mu_1) \\ \phi_{\mathrm{VIX}^2} \in L^1(\mu_{\mathrm{VIX}^2})}} \mathbb{E}^{\mu_1}[\phi_1] + \mathbb{E}^{\mu_{\mathrm{VIX}^2}}[\phi_{\mathrm{VIX}^2}]$$

is upper semi-continuous on $C_\psi(\mathbb{R}_{>0})$ equipped with the ψ-strict topology where we use

Lemma 5.6.6 to determine the necessary neighbourhoods in the topology. To this end by Lemma 5.6.6 choose $\phi \in C_0(\mathbb{R}_{>0})$ such that $\frac{1}{\phi\psi}$ has finite expectation under μ_2 for some μ_2 with $\mathcal{P}(\mu_1, \mu_{\mathrm{VIX}^2}, \mu_2) \neq \emptyset$. For any $f \in C_\psi(\mathbb{R}_{>0})$ set again

$$U_\epsilon(f) := \{g \in C_\psi(\mathbb{R}_{>0}) : \ |f - g|\phi\psi < \epsilon\},$$

which are clearly open neighbourhoods in the ψ-strict topology. We again have the bound

$$g \leq f + \frac{\epsilon}{\phi\psi}$$

for $g \in U_\epsilon(f)$ Assume that

$$F(f) = \sup_{\substack{\phi_1 + \phi_{\mathrm{VIX}^2} \leq S_P f \\ \phi_1 \in L^1(\mu_1) \\ \phi_{\mathrm{VIX}^2} \in L^1(\mu_{\mathrm{VIX}^2})}} \mathbb{E}^{\mu_1}[\phi_1] + \mathbb{E}^{\mu_{\mathrm{VIX}^2}}[\phi_{\mathrm{VIX}^2}] < r$$

for some $r \in \mathbb{R}$. For $g \in U_\epsilon(f)$ we obtain

$$\sup_{\substack{\phi_1 + \phi_{\mathrm{VIX}^2} \leq S_P g \\ \phi_1 \in L^1(\mu_1) \\ \phi_{\mathrm{VIX}^2} \in L^1(\mu_{\mathrm{VIX}^2})}} \mathbb{E}^{\mu_1}[\phi_1] + \mathbb{E}^{\mu_{\mathrm{VIX}^2}}[\phi_{\mathrm{VIX}^2}] = \inf_{\mu \in \mathcal{C}(\mu_1, \mu_{\mathrm{VIX}^2})} \mathbb{E}^\mu[S_P g] \leq \inf_{\mu \in \mathcal{C}(\mu_1, \mu_{\mathrm{VIX}^2})} \mathbb{E}^\mu[S_P(f + \frac{\epsilon}{\psi\phi})]$$

Since S_P is continuous from above, i.e. continuous for pointwise monotonically decreasing sequences, and monotone we have by monotone convergence

$$\inf_{\epsilon > 0} \inf_{\mu \in \mathcal{C}(\mu_1, \mu_{\mathrm{VIX}^2})} \mathbb{E}^\mu[S_P(f + \frac{\epsilon}{\psi\phi})] = \inf_{\mu \in \mathcal{C}(\mu_1, \mu_{\mathrm{VIX}^2})} \mathbb{E}^\mu[S_P f] = F(f) < r.$$

Hence there exists an $\epsilon > 0$ such that for all $g \in U_\epsilon(f)$ we have $F(g) < r$ and thus F is upper semi-continuous. Since F is superlinear and upper semi-continuous there exists a convex set $B \subset M_\psi(\mathbb{R}^d)$ such that we have

$$F(f) = \inf_{\lambda \in B} \mathbb{E}^\lambda[f].$$

Plugging in $f \equiv \pm 1$ we obtain that each $\lambda \in B$ has total mass 1. For all $f \in C_\psi(\mathbb{R}_{>0})$ with $f \geq 0$ we have

$$\inf_{\lambda \in B} \mathbb{E}^\lambda[f] = F(f) \geq \mathbb{E}^{\mu_1}[0] + \mathbb{E}^{\mu_{\mathrm{VIX}^2}}[0] = 0.$$

Hence all $\lambda \in B$ are probability measures and thus by the first part of the proof $\mathcal{P}(\mu_1, \mu_{\mathrm{VIX}^2}, \lambda) \neq \emptyset$ since each λ satisfies the existence conditions. Thus $B \subset \{\lambda : \mathcal{P}(\mu_1, \mu_{\mathrm{VIX}^2}, \lambda) \neq \emptyset\}$. But every $\mu_2 \in \{\lambda : \ \mathcal{P}(\mu_1, \mu_{\mathrm{VIX}^2}, \lambda) \neq \emptyset\}$ satisfies by Theorem

5.3.7

$$\sup_{\substack{\phi_1+\phi_{\mathrm{VIX}^2}\le S_P f\\ \phi_1\in L^1(\mu_1)\\ \phi_{\mathrm{VIX}^2}\in L^1(\mu_{\mathrm{VIX}^2})}} \mathbb{E}^{\mu_1}[\phi_1]+\mathbb{E}^{\mu_{\mathrm{VIX}^2}}[\phi_{\mathrm{VIX}^2}]\le \mathbb{E}^{\mu_2}[f]$$

and thus the reverse inclusion is obvious and we obtain equality. □

5.A. The convex conjugate

For a more complete overview of the convex conjugate, its properties and proofs for the statements below, see [64] for the finite dimensional case and [63] for the infinite dimensional case. Let X be a locally convex Hausdorff space and denote by X^* its topological dual equipped with the weak-$*$-topology. Further denote by $\langle\cdot,\cdot\rangle$ the dual pairing

$$\langle\cdot,\cdot\rangle : ((x^*,x)\in X^*\times X)\mapsto x^*(x).$$

Definition 5.A.1.

We say that an extended real function $f : X\to\overline{\mathbb{R}}$ is *proper* if and only if it does not attain the value $-\infty$ and is not constant ∞.

Definition 5.A.2. *For any function $f : X\to\overline{\mathbb{R}}$ define its convex conjugate $f^* : X^*\to\overline{\mathbb{R}}$ by*

$$f^*(x^*)=\sup_{x\in X}\langle x^*,x\rangle - f(x).$$

Lemma 5.A.3 (Fenchel's inequality)**.** *For all $x\in X$, $x^*\in X^*$ and all $f : X\to\overline{\mathbb{R}}$ we have*

$$\langle x^*,x\rangle\le f(x)+f^*(x^*).$$

Theorem 5.A.4 (Fenchel-Moreau)**.** *If $f : X\to\overline{\mathbb{R}}$ is*

- *constant ∞*
- *constant $-\infty$ or*
- *proper, lower semi-continuous and convex*

then

$$f^{**}:=(f^*)^*=f$$

where f^{**} is known as the biconjugate of f.

An important property of the biconjugate is that it preserves ordering. If $g : X\to\overline{\mathbb{R}}$ satisfies $g\le f$ then we also have $g^{**}\le f^{**}$.
Since for any $f : X\to\overline{\mathbb{R}}$ its convex conjugate is given by a supremum over affine continuous functionals, f^* is lower semi-continuous and convex on X^* and thus by the same argument the biconjugate f^{**} is lower semi-continuous and convex on X.

Definition 5.A.5 (Lower semi-continuous convex envelope)**.** *Let $f : X \to \overline{\mathbb{R}}$. The lower semi-continuous envelope of f denoted by $\mathcal{L}f : X \to \overline{\mathbb{R}}$ is given by*

$$\mathcal{L}f(x) := \sup\{g(x)|\ g \text{ convex and lower semi-continuous and } g \leq f\}.$$

$\mathcal{L}f$ is the unique maximal element in set of lower semi-continuous convex functions dominated by f.

Remark 5.A.6. *If X is finite dimensional, then any lower semi-continuous convex $f : X \to \overline{\mathbb{R}}$ has a closed and convex epigraph. Thus by the supporting hyperplane theorem the set of subdifferential is non-empty and we obtain the equivalent formulation*

$$\mathcal{L}f(x) := \sup\{g(x)|\ g \text{ affine and lower semi-continuous and } g \leq f\}.$$

The following corollary holds in a more generality, see Theorem 5 of [63]. For completeness, we provide the proof here under slightly more restrictive conditions.

Corollary 5.A.7. *For any function $f : X \to \overline{\mathbb{R}}$ such that either $f \equiv \infty$, it is proper and bounded below, then the biconjugate f^{**} is given by the lower semi-continuous convex envelope $\mathcal{L}f$.*

Proof. This is obvious for $f \equiv \infty$. Assume that f is proper and bounded below. Since f is bounded below, so is $\mathcal{L}f$. Clearly $\mathcal{L}f \leq f$ and so $\mathcal{L}f$ is proper. By Theorem 5.A.4 we have $(\mathcal{L}f)^{**} = \mathcal{L}f$. From Fenchel's inequality 5.A.3 we have

$$\langle x^*, x\rangle - f^*(x^*) \leq f(x)$$

and thus taking the supremum over $x^* \in X^*$ we obtain $f^{**} \leq f$. From $\mathcal{L}f \leq f$ we obtain $\mathcal{L}f = (\mathcal{L}f)^{**} \leq f^{**}$ and since f^{**} is lower semi-continuous and convex and is dominated by f we have $\mathcal{L}f \geq f$. □

Definition 5.A.8. *A map $p : X \to \overline{\mathbb{R}}$ is sublinear if for all $x, y \in X$ p is subadditive, i.e.*

$$p(x) + p(y) \leq p(x + y)$$

and positive homogeneous, i.e. for all $x \in X$ and $\alpha \geq 0$ we have $p(\alpha x) = \alpha p(x)$. A map $p : X \to \overline{\mathbb{R}}$ is superlinear if $x \mapsto -p(-x)$ is sublinear.

The next Theorem holds in more generality, see Theorem 2.4.14 in [76]

Theorem 5.A.9. *For every lower semi-continuous proper sublinear map $p : X \to \overline{\mathbb{R}}$ there exists a closed convex set $A_p \subset X^*$ such that p is given by*

$$p(x) = \sup_{x^* \in A_p} \langle x^*, x\rangle.$$

Conversely for every upper semi-continuous superlinear map $p : X \to \overline{\mathbb{R}}$ with $x \mapsto -p(x)$

proper there exists a closed convex set $A_p \subset X^*$ such that p is given by

$$p(x) = \inf_{x^* \in A_p} \langle x^*, x\rangle.$$

Proof. Assume p is sublinear. Clearly by subadditivity any sublinear map is convex. Hence by Theorem 5.A.4 we have $p^{**} = p$. We have

$$p^*(x^*) = \sup_{x \in X} \langle x^*, x\rangle - p(x).$$

If $p^*(x^*) \leq 0$ then by positive homogeneity $p^*(x^*) = 0$. Conversely if $p^*(x^*) > 0$ then again by positive homoegeneity we obtain $p^*(x^*) = \infty$. Set

$$A_p := \{x^* |\, p^*(x^*) = 0\} = \{x^* |\, p^*(x^*) \leq 0\}.$$

By convexity and lower semi-continuity of p^* the set A_p is closed and convex. By the above argument we have

$$p^*(x^*) = \begin{cases} 0 & x^* \in A_p \\ \infty & \text{else.} \end{cases}$$

By Theorem 5.A.4 we obtain for any $x \in X$

$$p(x) = \sup_{x^* \in X^*} \langle x^*, x\rangle - p^*(x^*) = \sup_{x^* \in A_p} \langle x^*, x\rangle.$$

Assume now that p is superlinear. By definition the map $q : x \mapsto -p(-x)$ is sublinear and thus we obtain

$$-p(x) = q(-x) = \sup_{x^* \in A_q} \langle x^*, -x\rangle.$$

and thus

$$p(x) = \inf_{x^* \in A_q} \langle x^*, x\rangle.$$

□

5.B. Measurability of S_P

Lemma 5.B.1. *Let $A \subset X$ be compact and let $f : X \to \overline{\mathbb{R}}$ be lower semi-continuous and $f > -\infty$, then $\overline{f}_A(x)$ is lower semi-continuous and $\overline{f}_A(x) > -\infty$.*

Proof. As A is compact and f is lower semi-continuous and $f > -\infty$, it follows that f is bounded from below on A. Clearly if $x \notin \mathrm{conv}(A)$ we have $\overline{f}_A(x) = \infty$. Hence assume $x \in \mathrm{conv}(A)$ and let $(x_n)_n$ be a sequence in A with $\overline{f}_A(x_n) < \infty$ converging to x. The set of probability measures μ with support in A and $|\,\mathrm{supp}(\mu)| \leq d + 1$ is weakly compact. Hence by Lemma 5.4.5 there exists $\mu_n \in \Xi_A^{x_n}(f)$ such that $\overline{f}_A(x_n) = \mathbb{E}^{\mu_n}[f]$. By compactness the sequence μ_n must have a cluster point μ with barycenter x and finite support. Thus

$$\liminf_{n\to\infty} \overline{f}_A(x_n) = \liminf_{n\to\infty} \mathbb{E}^{\mu_n}[f] \geq \mathbb{E}^{\mu}[f] \geq \overline{f}_A(x)$$

where the first inequality follows from f being lower semi-continuous. It follows that $\overline{f}_A$ is lower semi-continuous. There exists K such that $f(x) \geq K$ for $x \in A$. It follows that $\overline{f}_A(x) \geq K$ for all x. □

Lemma 5.B.2. *Let $A \subset X$ be compact, f lower semi-continuous and $A \subset \{f > -\infty\}$, then we have $\overline{f}_A(x) = (f_A)^{**}(x)$ (see Definition 5.B.3) and*

$$f_A(x) = \begin{cases} f(x) & x \in A, \\ \infty & \textit{else}. \end{cases}$$

Proof. Since f_A is bounded below by Corollary 5.A.7 $(f_A)^{**} = \mathcal{L}(f_A)$ is the largest convex lower semi-continuous function smaller than f_A. By Lemma 5.B.1 $\overline{f}_A$ is lower semi-continuous and convex. Furthermore $\overline{f}_A$ satisfies $\overline{f}_A \leq f_A$ since for any $x \in A$ with $f(x) < \infty$ we have by 5.4.3

$$f_A(x) = \mathbb{E}^{\delta_x}[f] \geq \inf_{\mu \in \Xi_A^x(f)} [f] = \overline{f}_A(x).$$

We have for any $x \in X$ that

$$\inf_{\mu \in \Xi_A^x(f)} \mathbb{E}^{\mu}[f] = \inf_{\mu \in \Xi_A^x(f)} \mathbb{E}^{\mu}[f_A]$$

since each measure $\mu \in \Xi_A^x(f)$ is supported on A. Hence

$$\begin{aligned} \overline{f}_A(x) &= \inf_{\mu \in \Xi_A^x(f)} \mathbb{E}^{\mu}[f] \\ &= \inf_{\mu \in \Xi_A^x(f)} \mathbb{E}^{\mu}[f_A] \\ &\geq \inf_{\mu \in \Xi_A^x(f)} \mathbb{E}^{\mu}[(f_A)^{**}] \geq (f_A)^{**}(x), \end{aligned}$$

where the last inequality follows from Jensen's inequality. Thus

$$f_A \geq \overline{f}_A \geq \mathcal{L}(f_A) = (f_A)^{**}$$

and thus by definition of the lower semi-continuous convex envelope we have equality. □

In the following let Y be a finite dimensional real vector space

Lemma 5.B.3. *Let $P \subset Y \times X$ be compact and $f : Y \times X \to \overline{\mathbb{R}}$ be lower semi-continuous and $f > -\infty$. Set*

$$g_y(x) = \begin{cases} f(y,x) & (y,x) \in P \\ \infty & \textit{else}, \end{cases}$$

then $(y,x) \mapsto (g_y)^{**}(x)$ *is jointly lower semi-continuous and does not attain* $-\infty$*.*

Proof. Since P is closed and f is lower semi-continuous, g is lower semi-continuous in (y,x). We have

$$(g_y)^*(x) = \sup_{z \in p_2(P)} \langle x, z\rangle_X - g_y(z),$$

where $p_2 : (y,x) \mapsto x$ and $\langle \cdot, \cdot \rangle_X$ denotes the standard inner product in X.
Notice that $(x,y,z) \mapsto \langle x, z\rangle - g_y(x)$ is jointly upper semi-continuous. As $p_2(P)$ is compact, since projections of compact sets are compact, we have by Proposition 7.32 in [15] that its supremum over a compact set with respect to z is also upper semi-continuous. Hence $(y,x) \mapsto (g_y)^*(x)$ is jointly upper semi-continuous. We have

$$(g_y)^{**}(x) = \sup_{z \in X} \langle x, z\rangle_X - (g_y)^*(z).$$

$(x,y,z) \mapsto \langle x, z\rangle - (g_y)^*(z)$ is lower semi-continuous and since the supremum of lower semi-continuous maps is again lower semi-continuous we obtain that $(y,x) \mapsto (g_y)^{**}(z)$ is (jointly) lower semi-continuous. Since f is lower semi-continuous it attains a minimum value on P. Hence, for all $x, y \in X \times Y$, $g_y(x) > C$ for some $C \in \mathbb{R}$. Since the biconjugate preserves ordering we obtain $(g_y)^{**}(x) > C$ and thus $(y,x) \mapsto (g_y)^{**}(x)$ does not attain $-\infty$. □

Definition 5.B.4. *For any* $P \subset Y \times X$ *set* $P_y := \{x \in X : \ (y,x) \in P\}$*.*

Lemma 5.B.5. *If* $P \subset Y \times X$ *is compact and* $f : Y \times X \to \overline{\mathbb{R}}$ *lower semi-continuous and does not attain* $-\infty$ *then* $Y \times X \ni (y,x) \mapsto \overline{f(y,\cdot)}_{P_y}(x)$ *is jointly lower semi-continuous and does not attain* $-\infty$*.*

Proof. Define

$$g_y(x) = \begin{cases} f(y,x) & (y,x) \in P \\ \infty & \text{else,} \end{cases}$$

then $g_y(x)$ is jointly lower semi-continuous and we have $g_y(x) = f(y,\cdot)_{P_y}(x)$. Hence by Lemma 5.B.3

$$(g_y)^{**}(x) = (f(y,\cdot)_{P_y})^{**}(x) = \overline{f(y,\cdot)}_{P_y}(x)$$

and $(g_y)^{**}(x)$ is jointly lower semi-continuous and greater than $-\infty$ by Lemma 5.B.3 □

Definition 5.B.6. *We say that* $f : X \to \overline{\mathbb{R}}$ *is* σ*-lower semi-continuous if there exists a pointwise decreasing sequence* $(f_n)_n$ *of lower semi-continuous functions converging pointwise to* f*.*

Lemma 5.B.7. *The class of* σ*-lower semi-continuous functions is closed under taking pointwise decreasing limits.*

Proof. Let $(f_n)_n$ be a decreasing sequence of σ-lower semi-continuous functions converging to f. For each n let f_n^m be a decreasing sequence of lower semi-continuous functions converging to f_n. We claim that $g_n = \min(f_1^n, \dots, f_n^n)$ converges pointwise to f. Clearly we have $g_n \geq f_n$ since $f_i^n \geq f_i$ and the sequence f_n is decreasing. Hence it follows that $\liminf_n g_n(x) \geq f(x)$. Conversely for any $n \geq m$ we have $g_n \leq f_m^n$ and thus for any m we have $\limsup_n g_n(x) \leq f_m(x)$ and hence $\limsup_n g_n(x) \leq \inf_n f_n(x) = f(x)$. □

Recall Definition 5.4.6.

Theorem 5.B.8. *Let $P \subset Y \times X$ be σ-compact and $f : Y \times X \to \overline{\mathbb{R}}$ be σ-lower semi-continuous. Then $H_P f$ is σ-lower semi-continuous.*

Proof. Let $(\widetilde{f}_n)$ be a decreasing sequence of lower semi-continuous function converging to f. Set $f_n = \max(\widetilde{f}_n, -n)$ and $(P_n)_n$ a sequence of increasing compact sets such that $\cup P_n = P$. By Lemma 5.B.5, $H_{P_n}(f_n)$ is lower semi-continuous. By Lemma 5.4.8 we have

$$\lim_{n\to\infty} H_{P_n}(f_n) = H_P f.$$

Hence $H_P f$ is σ-lower semi-continuous. □

Theorem 5.B.9 (Measurability of S_P)**.** *For any σ-lower semi-continuous function $f : X \to \overline{\mathbb{R}}$ the function $S_P f$ is σ-lower semi-continuous and hence measurable.*

Proof. We have that $f \circ p_n : X^n \to \overline{\mathbb{R}}$ is also σ-lower semi-continuous and hence by Theorem 5.B.8 applied to $Y = X^{n-1}$ we obtain that $S_P^n(f \circ p_n)$ is measurable. Proceeding inductively and reapplying Theorem 5.B.8 we obtain that

$$S_P^i \circ \dots S_P^n(f \circ p_n) : X^{i-1} \to \overline{\mathbb{R}}$$

is σ-lower semi-continuous. Thus $S_P f$ is σ-lower semi-continuous as well and hence measurable. □

5.C. Auxiliary results

In this section we recall some known results.

Lemma 5.C.1. *Let $X = \mathbb{R}^d$ and let $A \subset X$ be nonempty. Then we have*

$$\{\pi(\lambda) : \ \lambda \in \Pi_A\} = \operatorname{conv}(A)$$

where conv(A) *is the convex hull of A.*

Proof. Without loss of generality assume that the affine hull of A is X as otherwise we could restrict ourselves to a subspace of X of lower dimension. Let $x \in \operatorname{conv}(A)$ then there is $n \in \mathbb{N}$, $\alpha_1, \dots \alpha_n \geq 0$ with $\sum_{i=1}^n \alpha_i = 1$ and $x_1, \dots, x_n \in A$ such that $\sum_{i=1}^n \alpha_i x_i = x$. Set $\lambda_x = \sum_{i=1}^n \alpha_i \delta_{x_i}$ where δ is the Dirac measure. Then $\lambda_x \in \Pi_A$ and $\pi(\lambda(x)) = x$. Hence we have $\operatorname{conv}(A) \subset \{\pi(\lambda) : \ \lambda \in \Pi_A\}$.
We show the reverse inclusion with induction on the dimension d. For $d = 0$ it is trivial.

Denote by $\operatorname{cl} B$ the closure of $B \subset X$ in X. Let $\lambda \in \Pi_A$. Assume $\pi(\lambda) \notin \operatorname{cl}(\operatorname{conv}(A))$ then $\{\pi(\lambda)\}$ and $\operatorname{cl}(\operatorname{conv}(A))$ can be strictly separated, i.e. there exists $y \in \mathbb{R}^d$ such that

$$\langle y, \pi(\lambda)\rangle < \inf_{x \in \operatorname{cl}(\operatorname{conv}(A))} \langle y, x\rangle.$$

Since λ is supported on A we have

$$\langle y, \pi(\lambda)\rangle = \langle y, \int x d\lambda\rangle = \int \langle y, x\rangle d\lambda > \int \langle y, \pi(\lambda)\rangle d\lambda = \langle y, \pi(\lambda)\rangle$$

which is a contradiction. Hence $\pi(\lambda) \in \operatorname{cl}(\operatorname{conv}(A))$. Denote by $(\cdot)^\circ$ the interior of a set. Since the affine hull of A is X the relative interior of $\operatorname{conv}(A)$ is equal to the interior $\operatorname{conv}(A)^\circ$ and in particular it is nonempty. For any convex set $C \subset X$ with $C^\circ \neq \emptyset$ we have $\operatorname{cl} C^\circ = C^\circ$. Assume $\pi(\lambda) \in \partial \operatorname{cl}(\operatorname{conv}(A))$, where $\partial(\cdot)$ denotes the boundary of a set, then, by the supporting hyperplane theorem there exists $y \in X$ such that

$$\langle y, \pi(\lambda)\rangle = \inf_{x \in \operatorname{cl}(\operatorname{conv}(A))} \langle y, x\rangle.$$

Hence for $x \in \operatorname{supp}(\lambda)$ we have $\langle y, x - \pi(\lambda)\rangle \geq 0$. We obtain

$$0 \leq \int \langle y, x - \pi(\lambda)\rangle d\lambda = \langle y, \int x d\lambda - \pi(\lambda)\rangle = 0$$

and thus $\operatorname{supp}(\lambda) \subset H$ where $H = \{x : \langle y, x - \pi(\lambda)\rangle = 0\}$. It follows that $\operatorname{supp}(\lambda) \subset H \cap A$. Since H is of lower dimension and hence by induction we obtain $\pi(\lambda) \in \operatorname{conv}(H \cap A) \subset \operatorname{conv}(A)$. If $\pi(\lambda) \notin \partial \operatorname{cl}(\operatorname{conv}(A))$, then $\pi(\lambda) \in \operatorname{conv}(A)^\circ \subset \operatorname{conv}(A)$ and we are done. □

Since convex functions are classically defined through an inequality condition, we will now show that the convex functions defined via the convexity of their epigraphs, as in [64], also satisfy this classical definition.

Lemma 5.C.2. *Let $f : X \to \overline{\mathbb{R}}$ be convex. Let $x_1, \ldots, x_n \in X$ such that $f(x_i) < \infty$. Then all $\alpha_1, \ldots, \alpha_n \in [0,1]$ with $\sum_i \alpha_i = 1$ we have*

$$\sum_i \alpha_i f(x_i) \geq f\left(\sum_i \alpha_i x_i\right)$$

Proof. For all t_i, such that $(x_i, t_i) \in \operatorname{epi}(f)$ we have $t_i \geq f(x_i)$. Since $\operatorname{epi}(f)$ is convex we obtain $(\sum_i \alpha_i x_i, \sum_i \alpha_i t_i) \in \operatorname{epi}_S(f)$ and thus

$$\sum_i \alpha_i t_i \geq \left(\sum_i \alpha_i x_i\right)$$

since this holds for all $t_i \geq f(x_i)$ the result follows. □

In the following we provide a proof of Jensen's inequality for convex functions taking values in the extended reals, since the standard proofs usually require the functions to be real valued.

Lemma 5.C.3 (Jensen). *Let μ be a probability measure on X and let f be convex and $\mu(\{f < \infty\}) = 1$, then we have*

$$\mathbb{E}^{\mu}[f] \geq f(\pi(\mu)).$$

Proof. We use induction on the dimension d. Since $\operatorname{epi}(f)$ is convex, so is its projection $C := \{x : (x,t) \in \operatorname{epi}(f)\}$. By assumption we have $\mu(C) = 1$. W.l.o.g. assume that C has nonempty interior and $\pi(\mu)$ in C° since otherwise by Hahn-Banach we can restrict ourselves to a subspace H with $\mu(H \cap C) = 1$ and use the induction hypothesis. Assume there exists $x \in C^\circ$ such that $f(x) = -\infty$, then by convexity it follows that $f(y) = -\infty$ for all $y \in C^\circ$. This can seen by noticing that for any $y \in C^\circ$ there exists $\alpha > 1$ such that $x_0 = \alpha y + (1-\alpha)x \in C$. Let $(x_0, t_0) \in \operatorname{epi}(f)$ then for every $t \in \mathbb{R}$, $(x,t) \in \operatorname{epi}(f)$ and thus

$$\left(y, \frac{\alpha - 1}{\alpha} t + \frac{1}{\alpha} t_0\right) \in \operatorname{epi}(f).$$

Letting $t \to -\infty$ shows that $f(y) = -\infty$. Hence

$$-\infty = f(\pi(\mu)) \leq \mathbb{E}^{\mu}[f].$$

Else f is real valued on C°. It is well known that any real valued convex function on an open subset is continuous. It follows that f is continuous on C° and hence the set $\operatorname{epi}_S(f) \bigcap (C^\circ \times \mathbb{R})$ is open and convex where $\operatorname{epi}_S(f) = \{(x,t) : t \in \mathbb{R},\ t > f(x)\}$. By Hahn Banach there exists a linear functional l such that $l(x,t) > l(\pi(\mu), f(\pi(\mu)))$ for all $(x,t) \in \operatorname{epi}_S(f) \bigcap (C^\circ \times \mathbb{R})$. Hence there exists a linear function a such that we have

$$f(z) \geq f(\pi(\mu)) + a(z - \pi(\mu))$$

for $z \in C^\circ$. For $z \in \partial C$ there exists $t > 0$ such that $(1-t)\pi(\mu) + tz \in C^\circ$. We have

$$f(\pi(\mu)) + ta(z - \pi(\mu)) \leq f((1-t)\pi(\mu) + tz) \leq (1-t) f(\pi(\mu)) + t f(z)$$

where the second inequality follows from the convexity of f and the fact that $f < \infty$ on C. Hence we obtain

$$f(z) \geq f(\pi(\mu)) + a(z - \pi(\mu)).$$

We have

$$\mathbb{E}^{\mu}[f] \geq \mathbb{E}^{\mu}[f(\pi(\mu)) + a(\cdot - \pi(\mu))] = f(\pi(\mu)) + a(\pi(\mu) - \pi(\mu)) = f(\pi(\mu)).$$

$\square$

Bibliography

[1] *The CBOE volatility index-VIX*. https://www.cboe.com/tradable_products/vix/. Accessed: 2024-09-07.

[2] B. Acciaio, M. Beiglböck, F. Penkner, and W. Schachermayer, *A model-free version of the fundamental theorem of asset pricing and the super-replication theorem*, Mathematical Finance, 26 (2016), pp. 233–251.

[3] B. Acciaio, A. M. Cox, and M. Huesmann, *Model-independent pricing with insider information: A Skorokhod embedding approach*, Advances in Applied Probability, 53 (2021), pp. 30–56.

[4] J. Aczél, *Lectures on functional equations and their applications*, Mathematics in Science and Engineering, Vol. 19, Academic Press, New York-London, 1966.

[5] D. R. Adams and L. I. Hedberg, *Function spaces and potential theory*, vol. 314, Springer Science & Business Media, 2012.

[6] R. J. Adler, S. Cambanis, and G. Samorodnitsky, *On stable Markov processes*, Stochastic Process. Appl., 34 (1990), pp. 1–17.

[7] H. Alzer, *Converses of two inequalities of Ky Fan, O. Taussky, and J. Todd*, J. Math. Anal. Appl., 161 (1991), pp. 142–147.

[8] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer, *Existence, duality, and cyclical monotonicity for weak transport costs*, Calculus of Variations and Partial Differential Equations, 58 (2019), p. 203.

[9] J. Backhoff-Veraguas and G. Pammer, *Applications of weak transport theory*, Bernoulli, 28 (2022), pp. 370–394.

[10] C. Bayer, P. Friz, and J. Gatheral, *Pricing under rough volatility*, Quantitative Finance, 16 (2016), pp. 887–904.

[11] ——, *Pricing under rough volatility*, Quant. Finance, 16 (2016), pp. 887–904.

[12] C. Bayer, P. K. Friz, M. Fukasawa, J. Gatheral, A. Jacquier, and M. Rosenbaum, *Rough Volatility*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2023.

[13] M. Beiglböck, A. M. Cox, M. Huesmann, N. Perkowski, and D. J. Prömel, *Pathwise superreplication via Vovk's outer measure*, Finance and Stochastics, 21 (2017), pp. 1141–1166.

[14] M. Beiglböck, P. Henry-Labordere, and F. Penkner, *Model-independent bounds for option prices—a mass transport approach*, Finance and Stochastics, 17 (2013), pp. 477–501.

[15] D. Bertsekas and S. E. Shreve, *Stochastic optimal control: the discrete-time case*, vol. 5, Athena Scientific, 1996.

[16] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*, vol. 27 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1987.

[17] V. I. Bogachev and M. A. S. Ruas, *Measure theory*, vol. 2, Springer, 2007.

[18] T. Bojdecki, L. G. Gorostiza, and A. Talarczyk, *Some extensions of fractional Brownian motion and sub-fractional Brownian motion related to particle systems*, Electron. Comm. Probab., 12 (2007), pp. 161–172.

[19] I. S. Borisov, *A criterion for Gaussian random processes to be Markov processes*, Theory Probab. Appl., 27 (1983), pp. 863–865.

[20] R. C. Buck, *Bounded continuous functions on a locally compact space.*, Michigan Mathematical Journal, 5 (1958), pp. 95–104.

[21] Z. Chen, J. Fan, and K. Wang, *Multivariate Gaussian processes: definitions, examples and applications*, Metron, 81 (2023), pp. 181–191.

[22] P. Cheridito, M. Kiiski, D. J. Prömel, and H. M. Soner, *Martingale optimal transport duality*, Mathematische Annalen, 379 (2021), pp. 1685–1712.

[23] T. S. Chihara, *An introduction to orthogonal polynomials*, Mathematics and its Applications, Vol. 13, Gordon and Breach Science Publishers, New York-London-Paris, 1978.

[24] H. Cramér, *On the structure of purely non-deterministic stochastic processes*, Arkiv för matematik, 4 (1961), pp. 249–266.

[25] C. Cuchiero, G. Gazzani, J. Möller, and S. Svaluto-Ferro, *Joint calibration to SPX and VIX options with signature-based models*, Mathematical Finance, (2023).

[26] C. Cuchiero and J. Teichmann, *Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case*, Journal of evolution equations, 20 (2020), pp. 1301–1348.

[27] E. B. Davies, *One-parameter semigroups*, vol. 15 of London Mathematical Society Monographs, Academic Press, Inc., London-New York, 1980.

[28] N. G. de Bruijn, *Asymptotic methods in analysis*, vol. Vol. IV of Bibliotheca Mathematica [Mathematics Library], North-Holland Publishing Co., Amsterdam; P. Noordhoff Ltd., Groningen; Interscience Publishers Inc., New York, 1958.

[29] D. Dizdar and E. Kováč, *A simple proof of strong duality in the linear persuasion problem*, Games and Economic Behavior, 122 (2020), pp. 407–412.

[30] *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.23 of 2019-06-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

[31] Y. Dolinsky and H. M. Soner, *Martingale optimal transport in the Skorokhod space*, Stochastic Processes and their Applications, 125 (2015), pp. 3893–3931.

[32] J. L. Doob, *Stochastic processes*, John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1953.

[33] I. Eliazar, *Power Brownian motion*, J. Phys. A, 57 (2024), pp. 1–12, Paper No. 03LT01.

[34] P. Embrechts and M. Maejima, *Selfsimilar processes*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2002.

[35] K.-J. Engel and R. Nagel, *One-parameter semigroups for linear evolution equations*, vol. 194 of Graduate Texts in Mathematics, Springer-Verlag, New York, 2000.

[36] K. Fan, O. Taussky, and J. Todd, *Discrete analogs of inequalities of Wirtinger*, Monatsh. Math., 59 (1955), pp. 73–90.

[37] M. Forde, S. Gerhold, and B. Smith, *Small-time, large-time, and asymptotics for the Rough Heston model*, Mathematical Finance, 31 (2021), pp. 203–241.

[38] P. K. Friz, P. Gassiat, and P. Pigato, *Precise asymptotics: Robust stochastic volatility models*, (2021).

[39] J. Gatheral, T. Jaisson, and M. Rosenbaum, *Volatility is rough*, Quantitative finance, 18 (2018), pp. 933–949.

[40] J. Gatheral, P. Jusselin, and M. Rosenbaum, *The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem*, arXiv preprint arXiv:2001.01789, (2020).

[41] R. Goodman and N. R. Wallach, *Representations and invariants of the classical groups*, vol. 68 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1998.

[42] J. Guyon, *The joint S&P 500/VIX smile calibration puzzle solved*, Risk, April, (2020).

[43] T. Hida and M. Hitsuda, *Gaussian processes*, vol. 120, American Mathematical Soc., 1993.

[44] J. E. Humphreys, *Linear algebraic groups*, Graduate Texts in Mathematics, No. 21, Springer-Verlag, New York-Heidelberg, 1975.

[45] D. P. Huy, *A remark on non-Markov property of a fractional brownian motion*, Vietnam Journal of Mathematics, 31 (2003).

[46] D. P. Huy, *A remark on non-Markov property of a fractional Brownian motion*, Vietnam J. Math., 31 (2003), pp. 237–240.

[47] T. Ichiba, G. Pang, and M. S. Taqqu, *Path properties of a generalized fractional Brownian motion*, J. Theoret. Probab., 35 (2022), pp. 550–574.

[48] N. Jacobson, *Lie algebras*, Interscience Tracts in Pure and Applied Mathematics, No. 10, Interscience Publishers (a division of John Wiley & Sons, Inc.), New York-London, 1962.

[49] K. Karhunen, *Über die Struktur stationärer zufälliger Funktionen*, Arkiv för Matematik, 1 (1950), pp. 141–160.

[50] H. G. Kellerer, *Duality theorems for marginal problems*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 67 (1984), pp. 399–432.

[51] M. Kuczma and A. Zajtz, *On the form of real solutions of the matrix functional equation $\Phi(x)\Phi(y) = \Phi(xy)$ for non-singular matrices Φ*, Publ. Math. Debrecen, 13 (1966), pp. 257–262.

[52] M. Lifshits, *Lectures on Gaussian processes*, SpringerBriefs in Mathematics, Springer, Heidelberg, 2012.

[53] B. Lindström, *Determinants on semilattices*, Proc. Amer. Math. Soc., 20 (1969), pp. 207–208.

[54] G. Lunter, *New proofs and a generalization of inequalities of Fan, Taussky, and Todd*, J. Math. Anal. Appl., 185 (1994), pp. 464–476.

[55] C. B. Mehr and J. A. McFadden, *Certain properties of Gaussian processes and their first-passage times*, J. Roy. Statist. Soc. Ser. B, 27 (1965), pp. 505–522.

[56] G. V. Milovanović and I. v. Milovanović, *On discrete inequalities of Wirtinger's type*, J. Math. Anal. Appl., 88 (1982), pp. 378–387.

[57] T. Muir, *A treatise on the theory of determinants*, Dover Publications, Inc., New York, 1960. Revised and enlarged by William H. Metzler.

[58] I. Nourdin, *Selected aspects of fractional Brownian motion*, vol. 4 of Bocconi & Springer Series, Springer, Milan; Bocconi University Press, Milan, 2012.

[59] K. C. O'Meara, J. Clark, and C. I. Vinsonhaler, *Advanced topics in linear algebra*, Oxford University Press, Oxford, 2011.

[60] C. Pacati, G. Pompa, and R. Reno, *Smiling twice: the Heston++ model*, Journal of Banking & Finance, 96 (2018), pp. 185–206.

[61] G. Pang and M. S. Taqqu, *Nonstationary self-similar Gaussian processes as scaling limits of power-law shot noise processes and generalizations of fractional Brownian motion*, High Frequency, 2 (2019), pp. 95–112.

[62] R. M. Redheffer, *Easy proofs of hard inequalities*, in General inequalities, 3 (Oberwolfach, 1981), vol. 64 of Internat. Schriftenreihe Numer. Math., Birkhäuser, Basel, 1983, pp. 123–140.

[63] R. T. Rockafellar, *Conjugate duality and optimization*, SIAM, 1974.

[64] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1996.

[65] F. Russo and C. A. Tudor, *On bifractional Brownian motion*, Stochastic Process. Appl., 116 (2006), pp. 830–856.

[66] A. Sghir, *An extension of sub-fractional Brownian motion*, Publ. Mat., 57 (2013), pp. 497–508.

[67] ——, *A self-similar Gaussian process*, Random Oper. Stoch. Equ., 22 (2014), pp. 85–92.

[68] M. Sion, *On general minimax theorems.*, (1958).

[69] V. Skorniakov, *On a covariance structure of some subset of self-similar Gaussian processes*, Stochastic Process. Appl., 129 (2019), pp. 1903–1920.

[70] R. P. Stanley, *Enumerative combinatorics. Volume 1*, vol. 49 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, second ed., 2012.

[71] V. Strassen, *The existence of probability measures with given marginals*, The Annals of Mathematical Statistics, 36 (1965), pp. 423–439.

[72] C. Villani et al., *Optimal transport: old and new*, vol. 338, Springer, 2009.

[73] A. Yazigi, *Representation of self-similar Gaussian processes*, Statist. Probab. Lett., 99 (2015), pp. 94–100.

[74] ——, *Representations and regularity of Gaussian processes*, (2015).

[75] A. Zajtz, *On semigroups of linear operators*, Zeszyty Nauk. Uniw. Jagielloń. Prace Mat., (1971), pp. 181–184.

[76] C. Zalinescu, *Convex analysis in general vector spaces*, World Scientific Publishing Co., Inc, 2002.

[77] Z.-H. Zhang and Z.-G. Xiao, *A simple new proof of Fan-Taussky-Todd inequalities*, Aust. J. Math. Anal. Appl., 2 (2005), pp. 1–5.

[78] M. Zili, *Generalized fractional Brownian motion*, Mod. Stoch. Theory Appl., 4 (2017), pp. 15–24.