Dependence structures in Lévy-type Markov processes

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Eddie Brendan Tu

August 2017

© by Eddie Brendan Tu, 2017
All Rights Reserved.

I dedicate this dissertation to my parents, Albert and Amy Tu, and to my sister, Dr. Kelly Tu Frantz. My parents are my greatest inspiration. They faced war and communism and rose to the challenges as immigrants in America to give their children an opportunity for a good life. I am so grateful for the sacrifices they've given over the years and for the respect and humility they taught me. My sister Kelly is my greatest role model. She is a wizard in a constant juggling act of extraordinary challenges, which she always seems to overcome. Her passion and dedication to her field of study inspired me to obtain this degree.

Acknowledgments

I would like to begin by thanking my advisor Dr. Jan Rosinski for his patience and guidance over the last fours years. His mentorship changed how I think about mathematics and enabled me to overcome many challenges in my graduate career. Additionally, he is an inspiration as a teacher and presenter, and his willingness and ability to communicate difficult ideas, effectively and intuitively, spurred me to become an educator; for that, I am very grateful.

I would also like to thank my committee members Drs. Vasileios Maroulas, Yu-Ting Chen, and Haileab Hilafu. I am very grateful for the willingness of each of you to assist me, answer my questions, and guide me through this process of earning my Ph.D. In particular, I would like to thank Dr. Maroulas for constantly pushing us and challenging us in the classroom so that we could strive to be our best as mathematicians and presenters. A big thank you also goes out to Pam Armentrout for her calming presence and guidance over the past six years. I also want to thank Dr. Marie Jameson for her guidance in the process of helping me get a job and for the many fruitful conversations about teaching.

I want to thank the many friends and colleagues whom have been an amazing support system over the past six years: Brian Allen, Peter Jantsch, Tyler Massaro, Nate Pollesch, John Cummings, Kelly Rooker, Steve Fassino, Will Clagett, Marina Massaro, Nick Dexter, Gara Wolf, Joe Daws, Darrin Weber, Andrew Starnes, Andrew Marchese, Chase Worley, Ernest Jum, Kevin Sonnanburg, Greg Clark, and Liguo Wang.

I also want to thank my friends back in Virginia: Robbie, Matt, Dustin, Sam, Will, Pamela Palma, and Ray V., for being a constant reminder of keeping a good sense of humor and attitude in the face of difficult challenges. Lastly, and certainly not the least, a huge thanks goes out to my friend Raymond Wodarski. Your courage to overcome extreme difficulties and circumstances, day after day, gives me the strength to take on my challenges.

"[T]hat I could find company and consolation and hope in an object pulled almost at random from a bookshelf - felt akin to an instance of religious grace."

- Jonathan Franzen

Abstract

In this dissertation, we examine the positive and negative dependence of infinitely divisible distributions and Lévy-type Markov processes. Examples of infinitely divisible distributions include Poissonian distributions like compound Poisson and α-stable distributions. Examples of Lévy-type Markov processes include Lévy processes and Feller processes, which include a class of jump-diffusions, certain stochastic differential equations with Lévy noise, and subordinated Markov processes. Other examples of Lévy-type Markov processes are time-inhomogeneous Feller evolution systems (FES), which include additive processes. We will provide a tour of various forms of positive dependence, which include association, positive supermodular association (PSA), positive supermodular dependence (PSD), and positive orthant dependence (POD), and more. We will give a history of the characterization of these notions of positive dependence for infinitely divisible distributions, Lévy processes, and certain Feller diffusions. Additionally, we will present our contribution to the characterization of positive dependence for jump-Feller processes, and include applications. We will also characterize positive dependence for general time-inhomogeneous Feller evolution systems and jump-FESs. Finally, we characterize negative association and other forms of negative dependence for infinitely divisible distributions and Lévy processes.

Table of Contents

1 Introduction **1**

1.1 Notation 8

2 Positive dependence and infinite divisibility **10**

2.1 Association, weak association, and positive correlation 10

2.2 Positive supermodular association, supermodular dependence, orthant dependence 23

2.2.1 Stochastic orders 23

2.2.2 Dependence induced from stochastic orders 25

2.3 Positive dependence and infinite divisibility 37

2.3.1 Infinitely divisible (ID) distributions 37

2.3.2 Positive dependence in ID distributions 39

2.3.3 Positive dependence and infinite divisibility of squared Gaussians . . 45

3 Positive dependence in Lévy-type Markov processes **46**

3.1 Feller processes 46

3.1.1 Lévy processes 51

3.2 Positive dependence in general Markov processes 56

3.3 Positive dependence in Lévy processes 61

3.4 Positive dependence in Feller processes 61

3.4.1 Bounded symbols 62

3.4.2 Integro-differential operator, extended generator 63

3.4.3 Small-time asymptotics 65

3.4.4 Main results of this chapter . . . 66
3.5 Applications and examples . . . 86
3.5.1 Lévy processes . . . 86
3.5.2 Ornstein-Uhlenbeck process . . . 86
3.5.3 Feller's pseudo-Poisson process . . . 87
3.5.4 Bochner's subordination of a Feller process . . . 90
3.5.5 Lévy-driven stochastic differential equations . . . 92

4 Positive dependence in time-inhomogeneous Markov processes 98
4.1 Time-inhomogeneous Markov processes . . . 98
4.1.1 Time-homogeneous transformation of time-inhomogeneous Markov process . . . 101
4.2 Association of time-inhomogeneous Markov processes . . . 106
4.2.1 Temporal association . . . 113
4.2.2 Other forms of dependence in time-inhomogeneous Markov processes 114
4.3 Applications and examples . . . 116
4.3.1 Additive processes . . . 116
4.3.2 Comparison of Markov processes . . . 121

5 Negative dependence in ID distributions and Lévy processes 125
5.1 Various forms of negative dependence . . . 125
5.2 Negative dependence and infinite divisibility . . . 133
5.3 Negative dependence in Lévy processes . . . 142
5.4 Positive and negative dependence in limit theorems . . . 146

Bibliography 150

Vita 156

List of Figures

2.1 Implication map of positive dependence 29
2.2 Gaussian vs. α-stable 45

3.1 Equivalence of dependencies under condition (3.29) for Feller processes . . . 67
3.2 Transformed Lévy measure concentration 96

5.1 Implication map of negative dependence 131
5.2 Negative dependence for jump-Lévy processes 145

Chapter 1

Introduction

Association is an important property used to study positive dependence in multivariate distributions and stochastic processes. This property has a wide variety of applications, such as being a useful tool to measure system reliability in reliability theory and for proving limit theorems of interacting particle systems in physics. Additionally, association is useful in measuring ruin times of risk reserves in finance. This property also yields various central limit-type theorems for sequences of random vectors, random measures, and stochastic processes. Thus the characterization of association of multivariate distributions and stochastic processes is very important.

When a multivariate distribution, or random vector $X = (X_1, ..., X_d)$, exerts *positive correlation*, this is expressed mathematically as $\text{Cov}(X_i, X_j) \geq 0$ for all $i, j = 1, ..., d$, where Cov stands for the *covariance*. Oftentimes, however, it is desirable to have a stronger form of positive dependence, known as *association*.

Definition 1.1. A random vector $X = (X_1, ..., X_d)$ on $\mathbb{R}^d$ is **associated** if

$$\text{Cov}(f(X_1, ..., X_d), g(X_1, ..., X_d)) \geq 0$$

for every $f, g : \mathbb{R}^d \to \mathbb{R}$ non-decreasing in each component such that the covariance exists.

This is a stronger form of positive correlation. We can also talk about the association of stochastic processes in several different ways. The two ways we will discuss are given by the following definitions:

Definition 1.2. A stochastic process $X = (X_t)_{t\geq 0}$ on state space $\mathbb{R}^d$ is **associated in space** or **spatially associated** if the random vector $X_t = (X_t^{(1)}, ..., X_t^{(d)})$ is associated for all $t \geq 0$.

The intuitive meaning of this definition is that, as the process evolves, the strong positive dependence between the components $X_t^{(i)}$ is preserved.

Definition 1.3. A stochastic process $X = (X_t)_{t\geq 0}$ on $\mathbb{R}^d$ is **associated in time** or **temporally associated** if, for all $0 \leq t_1 < ... < t_n$, the finite-dimensional distribution $(X_{t_1}, ..., X_{t_n})$ is associated in $\mathbb{R}^{dn}$.

We can also look at these notions in a more general state space E, where E has partial ordering. However, for this dissertation, we will just consider the case $E = \mathbb{R}^d$. Temporal association is even stronger than Definition 1.2. Here, not only is there positive dependence between components for all t, but there is positive dependence between different moments in time of the process.

One can also look at weaker forms of positive dependence, such as positive supermodular association (PSA), positive supermodular dependence (PSD), and positive orthant dependence (POD), in multivariate distributions and stochastic processes. These are slightly weaker than association but still stronger than positive correlation.

If a stochastic process is a model for an evolving system, such as an interacting particle system in physics or a collection of risk reserves in finance, and the system exerts a strong positive dependence like association, PSA, PSD, or POD, either spatially (Def. 1.2) or temporally (Def. 1.3), then we would like to characterize the behavior of that process. This is the aim of the dissertation.

The stochastic processes in which we are interested are Markov processes with Lévy-type behavior. This behavior has an intimate connection with the distributional property known as **infinite divisibility (ID)**. Infinitely divisible distributions include a myriad of interesting examples, such as Gaussian distributions, Poisson and compound Poisson distributions, and α-stable distributions. Thus, to help us characterize the association of these Markov processes, we will first examine the characterization of association for infinitely divisible distributions by Pitt and Resnick [38], [39].

According to Lévy-Khintchine Theorem, an infinitely divisible random vector X has a nice decomposition into the sum of an independent Gaussian G and Poissonian P random vectors: $X = G + P$. Thus, the behavior of the ID distribution can be described by a characteristic (Lévy) triplet: (b, Σ, ν), where b, Q represent the mean and covariance, respectively, of the Gaussian G, and ν is a σ-finite measure representing the Poissonian behavior. Pitt (1982) used the parameter Σ to characterize the association of the Gaussian component: $\Sigma_{ij} \geq 0$ for all i, j [38]. Resnick (1988) used the Lévy measure ν to find a sufficient condition for association of the Poissonian component, with the condition being ν is concentrated on the positive and negative orthants: $\mathbb{R}^d_+ \cup \mathbb{R}^d_-$ [39].

A natural extension of the above results was to study association (in the sense of Defs. 1.2 and 1.3) for Lévy processes $X = (X_t)_{t \geq 0}$. For such processes, each X_t is infinitely divisible, and the process' behavior is described by a Lévy triplet (b, Σ, ν). More specifically, the process' behavior has an analogous decomposition to ID distributions, where process $X = G + P$, where G is a Brownian motion (with drift) and P is a Poissonian jump process. Hence, the Lévy triplet can be used to describe association in space and time. Association (time and space) of Brownian motion is the same as the characterization Pitt gave for Gaussian distributions. Samorodnitsky (1995) proved that the Lévy measure condition on the positive and negative orthants is sufficient and necessary condition for association in space for Poissonian jump processes [43]. Bauerle (2008) extended this notion to association in time [4].

Lévy processes are Markov processes which are homogeneous in space. However, many models of real-world phenomena do not exhibit such behavior. Thus it is important to study state space-dependent processes. We study Feller processes, a type of Markov process which is non-homogeneous in space, but has locally Lévy-type behavior. Feller processes arise as useful models in many applications in physics and finance (see [6]). The behavior of these processes can be described also by a Lévy triplet, except, unlike the Lévy case, the parameters are non-constant: $(b(x), \Sigma(x), \nu(x, dy))$, where x represents the starting location of the process in the state space. We can use this triplet to characterize the association of Feller processes as was done in the Lévy process case. Extending Pitt's result, Mu-Fa Chen (1993) characterized association for the case $(b(x), \Sigma(x), 0)$, i.e. diffusion processes [11]. Jie

Ming Wang (2009) characterized association for jump-Feller processes with characteristics $(b(x), 0, \nu(x, dy))$ under strong boundedness conditions and continuity conditions of the Lévy characteristics $b(x)$ and $\nu(x, dy)$ [52]. Our first contribution extends the work of Wang on jump-Feller processes by proving a similar characterization of association with relaxed continuity and integrability assumptions. This allows us to use the result for a larger class of Feller processes. Additionally, we characterize the stronger condition of association in time, not just association in space. Moreover, we prove results that characterize other forms of dependence, like PSA, PSD, and POD, in these jump-Feller processes. To prove this result, we do analysis on pseudo-differential operators, integro-differential operators, and small-time asymptotics of the Feller process. These contributions of ours can be found in [50].

Liggett (1985) [32] proved a useful characterization of the association of Markov processes using the generator of the process. The generator $\mathcal{A}$ generates the semigroup $(T_t)_{t\geq 0}$ of the process. The semigroup $(T_t)_{t\geq 0}$ is a family of bounded operators that describes the transition probabilities of the Markov process, while, intuitively, the generator $\mathcal{A}$ describes the rate of transition from one state to the next:

$$\mathcal{A} = \lim_{t \searrow 0} \frac{T_t - I}{t},$$

where the limit is taken in the uniform sense. Liggett's characterization was for Markov processes on a compact state space E with a bounded generator. For a **stochastically monotone** Markov process $X = (X_t)_{t\geq 0}$, i.e. $T_t f$ is non-decreasing if f non-decreasing for all $t \geq 0$, X will be spatially associated if and only if, for all $f, g \in C(E)$ non-decreasing,

$$\mathcal{A}fg \geq f\mathcal{A}g + g\mathcal{A}f. \tag{1.1}$$

Ruschendorf then extended this result to Markov processes on a Polish space E, with a possibly unbounded generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$, where $\mathcal{D}(\mathcal{A})$ represents the domain of generator $\mathcal{A}$ [42]. In the case of Lévy and Feller processes on $\mathbb{R}^d$, when one defines $\mathcal{A}$ as the uniform limit of $\frac{1}{t}(T_t - I)$, one normally works in the Banach space $(C_0(\mathbb{R}^d), ||\cdot||_\infty)$, where $||\cdot||_\infty$ represents the sup-norm. However, this would mean $\mathcal{F}_i \cap \mathcal{D}(\mathcal{A}) = \{0\}$, where $\mathcal{F}_i = \{f : \mathbb{R}^d \to \mathbb{R}, \text{ non-decreasing}\}$. Thus, for Lévy-type Markov processes, we would

like to extend the generator to a larger space of functions, such as $C_b(\mathbb{R}^d)$. In doing so, this extended generator may no longer generate the semigroups uniformly, but instead locally uniformly, or pointwise. The results of Schilling (1998) show that the integro-differential operator of the process is a natural extended generator of $\mathcal{A}$, defined on the space $C_b^2(\mathbb{R}^d)$ [45] [46]. The pseudo-differential and integro-differential operator is defined based on the **symbol** of the process: $p(x,\xi)$. The function $-p(x,\xi)$ is a continuous negative definite function in the second argument, for each x, and for a rich Feller process with Lévy triplet $(b(x), \Sigma(x), \nu(x,dy))$, the symbol $p(x,\xi)$ has representation

$$-p(x,\xi) = -ib(x)\cdot\xi + \frac{1}{2}\xi\cdot\Sigma(x)\xi - \int_{y\neq 0}(e^{iy\cdot\xi} - 1 - iy\cdot\xi\chi(y))\nu(x,dy) \tag{1.2}$$

where $\chi : \mathbb{R}^d \to \mathbb{R}$ is a cut-off function (Ex. $\chi(y) = \mathbb{1}_{B(0,1)}(y)$). By a result of Courrège [14], if $C_c^\infty(\mathbb{R}^d) \subset \mathcal{D}(\mathcal{A})$, where $C_c^\infty(\mathbb{R}^d)$ is the space of smooth functions with compact support on $\mathbb{R}^d$, then $-\mathcal{A}$ is a pseudo-differential operator $p(x,D)$ on $C_c^\infty(\mathbb{R}^d)$, i.e.

$$\mathcal{A}|_{C_c^\infty(\mathbb{R}^d)} f(x) = -p(x,D)f(x) := (2\pi)^{d/2}\int_{\mathbb{R}^d} e^{ix\cdot\xi}p(x,\xi)\hat{f}(u)du. \tag{1.3}$$

When we substitute (1.2) into the right-hand side of (1.3), the result is the **integro-differential operator** $I(p)$, given by

$$I(p)f(x) := b(x)\cdot\nabla f(x) + \frac{1}{2}\nabla\cdot\Sigma(x)\nabla f(x) + \int_{y\neq 0}(f(x+y) - f(x) - y\cdot\nabla f(x)\chi(y))\nu(x,dy) \tag{1.4}$$

The operator $I(p)$ has domain $C_b^2(\mathbb{R}^d)$, which, when intersected with $\mathcal{F}_i$, is a non-trivial class of functions, densely-defined in $C_b(\mathbb{R}^d) \cap \mathcal{F}_i$. To prove our main result for association of jump-Feller processes, we want to extend (1.1) to extended generators, specifically $I(p)$:

$$X_t \text{ associated } \forall t \geq 0 \iff I(p)fg \geq fI(p)g + gI(p)f \tag{1.5}$$

These results, along with characterizations of positive dependence of jump-Feller processes, and applications to certain stochastically monotone Feller processes will be presented in Chapter 3 and can be found in our paper [50].

Feller processes are time-homogeneous Markov processes. Certain real-world phenomena may not be time-homogeneous. Thus, it may be important to study time-inhomogeneous Markov processes. We examine a class of such processes, called Feller evolution systems (FES), which are strongly continuous, time-inhomogeneous Markov processes. These processes also have locally Lévy-type behavior. We extend our characterizations of positive dependence in Feller processes to these Feller evolutions systems. Our results are for general FES and jump-FESs. These results can be found in our paper [51].

Just as association is a strong form of positive correlation, negative association (NA) is a strong form of negative correlation.

Definition 1.4. A random vector $X = (X_1, ..., X_d)$ is **negatively associated (NA)** if, for all $I, J \subset \{1, ..., d\}$ disjoint, with $|I| = k$ and $|J| = n$, for all $f : \mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ non-decreasing in each component, we have

$$\mathrm{Cov}(f(X_I), g(X_J)) \leq 0$$

where $X_I := (X_i : i \in I)$ and $X_J := (X_j : j \in J)$. Essentially, X_I and X_J are disjoint sub-vectors of size k and n, respectively, of random vector X.

Negative association, in an analogous way to (positive) association and positive correlation, is a stronger form of negative correlation. Additionally, we can also study negative supermodular dependence (NSD) and negative orthant dependence (NOD), and other forms of negative dependence.

We study the NA of infinitely divisible distributions. The work on characterizing NA for Gaussian distributions has been done [27]. However, for Poissonian distributions, NA has not been characterized. Motivation for how to characterize NA comes from Lee et. al. [30], where the NA of α-stable random vectors is characterized. The characterization of NA of α-stable random vectors depends on the concentration of the *spectral measure* Γ of the distribution. This is a measure on the unit sphere S^d which describes the skewness and scale of the distribution. Because of its close connection to the Lévy measure, we can show a sufficient condition for NA of more general infinitely divisible Poissonian distributions. Analogous to positive dependence, we extend these results for NA of infinite divisibility

distributions to Lévy processes of the form $(b, 0, \nu)$. We prove that NA, NSD, NOD, and other forms of negative dependence are equivalent to the concentration of the Lévy measure given by (5.12). Our results on negative dependence can be found in a joint paper with Jan Rosinski [40], which will treat association and negative association of infinitely divisible distributions/processes and Lévy processes under a unified framework.

Finally, we give a brief overview of applications of association and negative association to limit theorems. The classical Laplace-de Moivre Central Limit Theorem looks at weak limits of rescaled sums of independent and identically distributed random variables, which converge to a Gaussian distribution:

$$\frac{X_1 + ... + X_n - n\mu}{\sigma n^{1/2}} \Rightarrow Z \sim N(0, I), \quad n \to \infty.$$

Similar results can be obtained for rescaled sums of associated and NA random variables. Additionally, one can have an analogous result to the Law of Large Numbers for associated and negatively associated random variables.

This dissertation is organized in the following way. In Chapter 2, we give an overview of different forms of positive dependence. We will also give an overview of infinitely divisible distributions and the characterization of association for those distributions. In Chapter 3, we examine the association of Lévy processes and Feller processes.. We will show our contribution to the characterization of positive dependence (association, PSA, PSD, POD, and more) for jump-Feller processes and give applications. In Chapter 4, we look at time-inhomogeneous Markov processes, specifically Feller evolution systems, and present our contribution to the characterization of positive dependence in those processes. In Chapter 5, we will discuss various forms of negative dependence and show our contribution to the characterization of negative dependence for infinitely divisible Poissonian distributions and jump-Lévy processes. We will also discuss some limit theorems of positively and negatively dependent sequences in this chapter.

1.1 Notation

$\mathbb{N} = \{1, 2, ...\}$ is the set of natural numbers, and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$ is the set of integers. $\mathbb{Z}_+ = \{0, 1, 2, ...\}$. $\mathbb{Q} = \{m/n, \text{ where } m, n \in \mathbb{Z}, \ n \neq 0\}$ is the set of rational numbers. $\mathbb{R}$ is the set of real numbers. $\mathbb{R}_+ = [0, \infty)$. $\mathbb{R}^{d\times d}$ is the space of $d \times d$ real matrices.

Throughout this dissertation, we are commonly in the setting of Euclidean space $\mathbb{R}^d$, where $d \in \mathbb{N}$. For elements $x = (x_1, ..., x_d)$ and $y = (y_1, ..., y_d)$ in $\mathbb{R}^d$, where $x_i, y_i \in \mathbb{R}$, we denote the inner product (or dot product) by

$$x \cdot y = \sum_{i=1}^{d} x_i y_i. \tag{1.6}$$

The corresponding Euclidean norm will be given by $|x| = (x \cdot x)^{1/2} = \left(\sum_{i=1}^{d} x_i^2\right)^{1/2}$. We will often deal with the following Banach spaces of functions: $B_b(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R}, \ ||f||_\infty < \infty\}$, where $||f||_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$. $C_b(\mathbb{R}^d)$ denotes the subspace of $B_b(\mathbb{R}^d)$ of continuous functions. $C_0(\mathbb{R}^d)$ denotes the subspace of $B_b(\mathbb{R}^d)$ of continuous functions that vanish at infinity, i.e. $|f(x)| \to 0$ as $|x| \to \infty$. $C_c(\mathbb{R}^d)$ denotes the subspace of $B_b(\mathbb{R}^d)$ of continuous functions with compact support. $C^n(\mathbb{R}^d)$ is the space of continuous functions such that all partial derivatives up to and including order n exist. $C^\infty(\mathbb{R}^d) = \bigcap_{n\in\mathbb{N}} C^n(\mathbb{R}^d)$. The Schwartz space $S(\mathbb{R}^d)$ is a subspace of $C^\infty(\mathbb{R}^d)$, where $f \in S(\mathbb{R}^d)$ satisfy $\sup_{x\in\mathbb{R}^d} |x^\alpha D^\beta f(x)| < \infty$, for all α, β multi-indices.

The triplet $(\Omega, \mathcal{G}, \mathbb{P})$ will often denote our probability space, where Ω is the sample space, $\mathcal{G}$ is the σ-algebra, and $\mathbb{P}$ is the probability measure. The state space of a random vector will often be in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. We say a random vector X has distribution μ, denoted $X \sim \mu$, where μ is a probability measure on $\mathbb{R}^d$, if $\mathbb{P}(X \in A) = \mu(A)$, for $A \in \mathcal{B}(\mathbb{R}^d)$.

The notation $\sim$ is also used to describe distributions of random vectors/variables. Bernoulli random variables are denoted $X \sim Bern(p)$, where $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$. Poisson random variables are denoted $X \sim Poiss(\lambda)$, with $\lambda > 0$, where $\mathbb{P}(X = n) = e^{-\lambda}\dfrac{\lambda^n}{n}$ for all $n \in \mathbb{Z}_+$. Gaussian random vectors are denoted $X \sim N(b, \Sigma)$, where $b \in \mathbb{R}^d$, and $\Sigma \in \mathbb{R}^{d\times d}$ is symmetric positive definite.

Oftentimes, the double right arrow $\Rightarrow$ will be used to denote weak convergence, i.e. $X_n \Rightarrow X$ means that X_n converges weakly (or converges in distribution) to X as $n \to \infty$. We often integrate over the set $\mathbb{R}^d \backslash \{0\}$. For simplicity, we will write $\int_{\mathbb{R}^d \backslash \{0\}}$ as $\int_{y \neq 0}$. Consider a random vector $X = (X_1, ..., X_d)$. Let $I \subset \{1, ..., d\}$. Then $X_I := (X_i : i \in I)$. For example, $I = \{s_1, ..., s_n\}$, $n \leq d$. Then $X_I = (X_{s_1}, ..., X_{s_n})$. The function $\chi : \mathbb{R}^d \to \mathbb{R}$ will likely always represent the cut-off function in the Lévy-Khintchine representation. Unless stated otherwise, we will always set $\chi(y) := \mathbb{1}_{(0,1)}(|y|)$.

Chapter 2

Positive dependence and infinite divisibility

In this chapter, we introduce different forms of positive dependence in multivariate distributions that are stronger than the usual characterization of "positive correlation." We will also introduce different stochastic orderings from which certain forms of positive dependence can be defined. These notions will be discussed, and important properties and examples will be provided. Our standard references for this topic are [49], [34], and [16],

2.1 Association, weak association, and positive correlation

Covariance is a standard tool to measure dependence in probability and statistics. For random variables X and Y, we define covariance to be

$$\mathrm{Cov}(X,Y) := \mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

$\mathrm{Cov}(X,Y) > 0$ indicates positive correlation, $\mathrm{Cov}(X,Y) < 0$ indicates negative correlation, and $\mathrm{Cov}(X,Y) = 0$ indicates no correlation between X and Y. We aim to study positive dependence in multivariate distributions $X = (X_1, ..., X_d)$. We say random vector X is **positively correlated (PC)** if $\mathrm{Cov}(X_i, X_j) \geq 0$ for all $i, j \in \{1, ..., d\}$.

It is known that, if $X = (X_1, ..., X_d)$ is positively correlated, positive dependence between components X_i is maintained under non-decreasing (or non-increasing) linear transformations $f = (f_1, ..., f_k) : \mathbb{R}^d \to \mathbb{R}^k$. f being linear means $f_i : \mathbb{R}^d \to \mathbb{R}$ is defined to be $f_i(x_1, ..., x_d) = a_1^{(i)} x_1 + ... + a_d^{(i)} x_d$ for all $i \in \{1, ..., k\}$, where coefficients $a_j^{(i)}$ are all non-negative (or all non-positive, respectively). In other words, if X is positively correlated in $\mathbb{R}^d$, then $f(X) = (f_1(X), ..., f_k(X))$ is also positively correlated in $\mathbb{R}^k$:

$$\begin{aligned} \mathrm{Cov}(f_i(X), f_j(X)) &= \mathrm{Cov}\left(\sum_{l=1}^{d} a_l^{(i)} X_l, \sum_{k=1}^{d} a_k^{(j)} X_k \right) \\ &= \sum_{l=1}^{d} \sum_{k=1}^{d} a_l^{(i)} a_k^{(j)} \mathrm{Cov}(X_l, X_k) \\ &\geq 0. \end{aligned}$$

Positive dependence will not be preserved under non-linear monotone transformation if X is just positively correlated, i.e. if X positively correlated, f non-linear non-decreasing function, then $f(X)$ may not be positively correlated. We provide a specific counter-example at the end of this section (see Example 2.1.1).

This shortcoming of positive correlation leads probabilists to consider stronger forms of positive dependence in which the positive dependence between components is preserved under non-linear monotone transformations. The following definition was first given by Esary, Proschan, and Walkup in 1967 [16].

Definition 2.1. A random vector $X = (X_1, ..., X_d)$ in $\mathbb{R}^d$ is **associated (A)** if

$$\mathrm{Cov}(f(X), g(X)) = \mathrm{Cov}(f(X_1, ..., X_d), g(X_1, ..., X_d)) \geq 0 \tag{2.1}$$

or equivalently

$$\mathbb{E} f(X) g(X) \geq \mathbb{E} f(X) \cdot \mathbb{E} g(X) \tag{2.2}$$

for all $f, g : \mathbb{R}^d \to \mathbb{R}$ non-decreasing in each component provided $\mathrm{Cov}(f(X), g(X))$ exists.

$f : \mathbb{R}^d \to \mathbb{R}$ **non-decreasing in each component** means $f_i(y) := f(x_1, ..., x_{i-1}, y, x_{i+1}, ..., x_d)$ is a non-decreasing function for all fixed x_k, $k \neq i$. Note that Definition 2.1 can also be stated

for all f, g non-increasing. Association is a stronger form of positive dependence than positive correlation. In other words:

Claim 2.1.1. *If $X = (X_1, ..., X_d)$ is associated and $\mathrm{Cov}(X_i, X_j) < \infty$ for all i, j, then X is positively correlated.*

Proof. Choose $f(x_1, ..., x_d) := x_i$ and $g(x_1, ..., x_d) := x_j$, for some $i, j \in \{1, ..., d\}$. Then f, g are both non-decreasing and

$$\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(f(X), g(X)) \geq 0.$$

□

Unlike the notion of positive correlation, association does preserve positive dependence between components under non-linear monotone transformations. In addition to this property, we get a collection of very nice properties.

Proposition 2.1.1. *Let $X = (X_1, ..., X_d)$ be a random vector.*

(a) Let $f = (f_1, ..., f_k) : \mathbb{R}^d \to \mathbb{R}^k$ be a non-decreasing function, i.e. $f_i : \mathbb{R}^d \to \mathbb{R}$ is non-decreasing in each component. If X is associated in $\mathbb{R}^d$, then $f(X)$ is associated in $\mathbb{R}^k$.

(b) Association is equivalent to the inequality

$$\mathbb{P}(X \in A \cap B) \geq \mathbb{P}(X \in A)\mathbb{P}(X \in B) \tag{2.3}$$

*for all monotone sets $A, B \in \mathcal{B}(\mathbb{R}^d)$. We say $A \in \mathcal{B}(\mathbb{R}^d)$ is **monotone** if $x \in A$, $x \leq y$ (componentwise: $x_i \leq y_i$, $\forall i$), then $y \in A$. Inequality (2.3) is equivalent to saying (2.1) holds for $f, g : \mathbb{R}^d \to \mathbb{R}$ non-decreasing indicator functions.*

(c) Association is equivalent to the inequality

$$\mathrm{Cov}(f(X), g(X)) \geq 0 \tag{2.4}$$

for all $f, g \in C_b(\mathbb{R}^d)$ non-decreasing.

(d) *Association is preserved under weak convergence, i.e. for a sequence of random vectors* $(X_n)_{n\in\mathbb{N}}$*, if* $X_n \Rightarrow X$ *and* X_n *is associated for all* n*, then* X *is associated.*

(e) *If* $X = (X_1, ..., X_d)$ *is associated in* $\mathbb{R}^d$ *and* $Y = (Y_1, ..., Y_k)$ *is associated in* $\mathbb{R}^k$*, and* X *and* Y *are independent, then* (X, Y) *is associated in* $\mathbb{R}^{d+k}$*. If* $d = k$*, then* $X + Y$ *is associated.*

(f) X *is always associated if* $d = 1$*.*

(g) *If* X *has independent components, then* X *is associated.*

Proof. Some of the ideas of the proofs of these properties can be found in [39].

(a) Let $f = (f_1, ..., f_k) : \mathbb{R}^d \to \mathbb{R}^k$ be a non-decreasing function. Then $g, h : \mathbb{R}^k \to \mathbb{R}$, non-decreasing. Then $g \circ f, h \circ f : \mathbb{R}^d \to \mathbb{R}$ are non-decreasing. By association of X,

$$\mathrm{Cov}(g(f(X)), h(f(X))) = \mathrm{Cov}((g \circ f)(X), (h \circ f)(X)) \geq 0.$$

The same result holds if $f : \mathbb{R}^d \to \mathbb{R}^k$ is non-increasing.

(b) Let $f, g : \mathbb{R}^d \to \mathbb{R}$ non-decreasing. Then by a formula of Lehmann [31], we have

$$\begin{aligned}
\mathrm{Cov}(f(X), g(X)) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{Cov}(\mathbb{1}_{(x,\infty)}(f(X)), \mathbb{1}_{(y,\infty)}(g(X))\, dx\, dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{Cov}(\mathbb{1}_{\{z\in\mathbb{R}^d : f(z)>x\}}(X), \mathbb{1}_{\{z\in\mathbb{R}^d : g(z)>y\}}(X)\, dx\, dy \\
&\geq 0
\end{aligned}$$

because $\{z \in \mathbb{R}^d : f(z) > x\}$ and $\{z \in \mathbb{R}^d : g(z) > y\}$ are monotone sets, and (2.3) implies the integrand is non-negative.

(c) Let $f, g : \mathbb{R}^d \to \mathbb{R}$, non-decreasing, such that $\mathrm{Cov}(f(X), g(X)) < \infty$. Define for $M \in \mathbb{N}$, the function $\tau_M : \mathbb{R} \to \mathbb{R}$, by

$$\tau_M(x) = \begin{cases} x & \text{if } |x| \leq M \\ M & \text{if } x > M \\ -M & \text{if } x < -M. \end{cases}$$

For all $M \in \mathbb{N}$, $\tau_M(f)$ is bounded, non-decreasing, and $\tau_M(f) \to f$ and $\tau_M(g) \to g$ pointwise as $M \to \infty$. Now let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be the standard Gaussian density function, i.e. $\varphi(x) = (2\pi)^{-d/2} \exp(-|x|^2/2)$. Define $\varphi_\epsilon(x) = \epsilon^{-d}\varphi(x/\epsilon)$ for all $\epsilon > 0$. Now let $h \in B_b(\mathbb{R}^d)$ and non-decreasing. Then define $h_\epsilon = h * \varphi_\epsilon$, i.e.

$$h_\epsilon(x) = \int_{\mathbb{R}^d} \varphi_\epsilon(x-y)h(y)dy = \int_{\mathbb{R}^d} h(x-y)\varphi_\epsilon(y)dy \tag{2.5}$$

Using the first integral representation, h_ϵ is bounded by h since the Gaussian density integrates to 1. Using the second integral representation, h_ϵ is continuous, by dominated convergence, and non-decreasing. (Additionally, one can also show $h_\epsilon \in C^\infty(\mathbb{R}^d)$). Finally, $h_\epsilon \to h$ pointwise, as $\epsilon \to 0$. Now define $f_{M,\epsilon}, g_{M,\epsilon} : \mathbb{R}^d \to \mathbb{R}$ by $f_{M,\epsilon} = \tau_M(f) * \varphi_\epsilon$ and $g_{M,\epsilon} = \tau_M(g) * \varphi_\epsilon$. Then $f_{M,\epsilon}, g_{M,\epsilon} \in C_b(\mathbb{R}^d)$, non-decreasing, for all $M \in \mathbb{N}$, $\epsilon > 0$. Then by assumption, $\mathbb{E}f_{M,\epsilon}(X)g_{M,\epsilon}(X) \geq \mathbb{E}f_{M,\epsilon}(X)\mathbb{E}g_{M,\epsilon}(X)$. By dominated convergence theorem,

$$\begin{aligned}
\mathbb{E}f(X)g(X) &= \lim_{M\to\infty} \mathbb{E}[\tau_M(f(X))\tau_M(g(X))] \\
&= \lim_{M\to\infty} \lim_{\epsilon\to 0} \mathbb{E}[(\tau_M(f) * \varphi_\epsilon)(X)(\tau_M(g) * \varphi_\epsilon)(X)] \\
&= \lim_{M\to\infty} \lim_{\epsilon\to 0} \mathbb{E}[f_{M,\epsilon}(X)g_{M,\epsilon}(X)] \\
&\geq \lim_{M\to\infty} \lim_{\epsilon\to 0} \mathbb{E}[f_{M,\epsilon}(X)] \cdot \mathbb{E}[g_{M,\epsilon}(X)] \\
&= \lim_{M\to\infty} \mathbb{E}[\tau_M(f(X))] \cdot \mathbb{E}[\tau_M(g(X))] \\
&= \mathbb{E}f(X)\mathbb{E}g(X)
\end{aligned}$$

(Note: $f_{M,\epsilon}, g_{M,\epsilon} \in C_b^\infty(\mathbb{R}^d)$, non-decreasing.)

(d) Let $f, g \in C_b(\mathbb{R}^d)$ non-decreasing. By weak convergence, dominated convergence, and part (c),

$$\begin{aligned}\mathbb{E}f(X)g(X) &= \lim_{n\to\infty} \mathbb{E}f(X_n)g(X_n) \\ &\geq \lim_{n\to\infty} \mathbb{E}f(X_n)\mathbb{E}g(X_n) \\ &= \mathbb{E}f(X)\mathbb{E}g(X)\end{aligned}$$

(e) Let $f, g : \mathbb{R}^{d+k} \to \mathbb{R}$ be non-decreasing, and $X \sim \mu_X$, $Y \sim \mu_Y$, where μ_X and μ_Y are the respective probability measures. Then by independence

$$\begin{aligned}&\mathrm{Cov}(f(X,Y), g(X,Y)) \\ &= \mathbb{E}fg(X,Y) - \mathbb{E}f(X,Y)\mathbb{E}g(X,Y) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^k} f(x,y)g(x,y)\mu_Y(dy)\mu_X(dx) \\ &\quad - \int_{\mathbb{R}^d}\int_{\mathbb{R}^k} f(x,y)\mu_Y(dy)\int_{\mathbb{R}^k} g(x,y)\mu_Y(dy)\mu_X(dx) \\ &\quad + \int_{\mathbb{R}^d}\int_{\mathbb{R}^k} f(x,y)\mu_Y(dy)\int_{\mathbb{R}^k} g(x,y)\mu_Y(dy)\mu_X(dx) \\ &\quad - \int_{\mathbb{R}^d}\int_{\mathbb{R}^k} f(x,y)\mu_Y(dy)\mu_X(dx)\cdot\int_{\mathbb{R}^d}\int_{\mathbb{R}^k} g(x,y)\mu_Y(dy)\mu_X(dx) \\ &= \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^k} f(x,y)g(x,y)\mu_Y(dy) - \int_{\mathbb{R}^k} f(x,y)\mu_Y(dy)\int_{\mathbb{R}^k} g(x,y)\mu_Y(dy)\right]\mu_X(dx) \\ &\quad + \mathbb{E}_X\left(\int_{\mathbb{R}^k} f(X,y)\mu_Y(dy)\int_{\mathbb{R}^k} g(X,y)\mu_Y(dy)\right) \\ &\quad - \mathbb{E}_X\left(\int_{\mathbb{R}^k} f(X,y)\mu_Y(dy)\right)\mathbb{E}_X\left(\int_{\mathbb{R}^k} g(X,y)\mu_Y(dy)\right) \\ &= \int_{\mathbb{R}^d}\mathrm{Cov}(f(x,Y), g(x,Y))\mu_X(dx) \\ &\quad + \mathrm{Cov}\left(\int_{\mathbb{R}^k} f(X,y)\mu_Y(dy), \int_{\mathbb{R}^k} g(X,y)\mu_Y(dy)\right),\end{aligned}$$

where the first term is non-negative since Y is associated, making the integrand non-negative, and the second term is non-negative, because X is associated and $\int_{\mathbb{R}^k} f(\cdot, y)\mu_Y(dy)$ is a non-decreasing function, since f is non-decreasing (similar for g). Hence, we have our desired result.

Now let $d = k$. For $f, g : \mathbb{R}^d \to \mathbb{R}$ non-decreasing

$$\begin{aligned}
\mathbb{E}f(X+Y)g(X+Y) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x+y)g(x+y)\mu_X(dx)\mu_Y(dy) \\
&= \int_{\mathbb{R}^d} \mathbb{E}_X f(X+y)g(X+y)\mu_Y(dy) \\
&\geq \int_{\mathbb{R}^d} \mathbb{E}_X f(X+y)\mathbb{E}_X g(X+y)\mu_Y(dy) \\
&= \mathbb{E}_Y[\mathbb{E}_X f(X+Y)\mathbb{E}_X g(X+Y)] \\
&\geq \mathbb{E}_Y\mathbb{E}_X f(X+Y)\cdot \mathbb{E}_Y\mathbb{E}_X g(X+Y)] \\
&= \mathbb{E}f(X+Y)\mathbb{E}g(X+Y)
\end{aligned}$$

where the first inequality comes from association of X, and the second inequality comes from association of Y and monotonicity of the expectation $\mathbb{E}_X$.

(f) Let A, B be monotone sets in $\mathbb{R}$. Then $A = (a, \infty)$ and $B = (b, \infty)$. WLOG, say $a < b$, which would make $A \cap B = B$. Thus,

$$\begin{aligned}
\mathbb{P}(X \in A \cap B) - \mathbb{P}(X \in A)\mathbb{P}(X \in B) &= \mathbb{P}(X \in B) - \mathbb{P}(X \in A)\mathbb{P}(X \in B) \\
&= \mathbb{P}(X \in B)[1 - \mathbb{P}(X \in A)] \\
&\geq 0
\end{aligned}$$

Thus, X is associated by part (b).

(g) We prove this by induction. Base Case: Let $d = 2$, since both X_1 and X_2 are associated in $\mathbb{R}$ by part (f), we have by part (e) that (X_1, X_2) is associated in $\mathbb{R}^2$. Induction Hypothesis: Now assume $(X_1, ..., X_{d-1})$ is associated with independent components. Inductive Step: Let X_d be independent of X_i, $i \in \{1, ..., d-1\}$. By (f) and (e), we have that $(X_1, ..., X_{d-1}, X_d)$ is associated.

□

Remark 2.1.1. As we mentioned earlier, the definition of association originally came from Esary et. al. in 1967 [16]. However, the idea was studied independently by Harris in 1960 and

Fortuin, Kastelyn, and Ginibre in 1971 from the perspective of percolation and statistical mechanics. In related literature in these areas, the notion was commonly referred to as "FKG inequalities," instead of "association." For more on this, see [20], [21].

One can also consider another strong form of positive dependence called weak association. It is weaker than the notion of association, but stronger than positive correlation.

Definition 2.2. A random vector $X = (X_1, ..., X_d)$ is **weakly associated (WA)** if, for any pair of disjoint subsets $I, J \subseteq \{1, .., d\}$, with $|I| = k$, $|J| = n$,

$$\text{Cov}(f(X_I), g(X_J)) \geq 0, \tag{2.6}$$

where $X_I := (X_i : i \in I)$, $X_J := (X_j : j \in J)$, for any $f : \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$ non-decreasing,

It is easy to see that WA is stronger than positive correlation.

Claim 2.1.2. *If $X = (X_1, ..., X_d)$ is weakly associated, then X is PC.*

Proof. Choose $i, j \in \{1, ..., d\}$, where $i \neq j$. Choose $I = \{i\}$, $J = \{j\}$. Then I, J disjoint, and now choose $f, g : \mathbb{R} \to \mathbb{R}$ to be the identity functions. By WA,

$$\begin{aligned} 0 &\leq \text{Cov}(f(X_I), g(X_J)) \\ &= \text{Cov}(f(X_i), g(X_j)) \\ &= \text{Cov}(X_i, X_j) \end{aligned}$$

□

Furthermore, WA is weaker than association.

Claim 2.1.3. *If $X = (X_1, ..., X_d)$ is associated, then X is weakly associated.*

Proof. Let $I, J \subseteq \{1, ..., d\}$ be disjoint subsets, with $|I| = k$, $|J| = n$. Define $f : \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$ to be non-decreasing. Then there exists $h, k : \mathbb{R}^d \to \mathbb{R}$ non-decreasing such that

$h(x_1, ..., x_d) := f(x_I)$ and $k(x_1, .., x_d) := g(x_J)$, where $x_I = (x_i : i \in I)$, $x_J = (x_j : j \in J)$. Then

$$0 \leq \text{Cov}(h(X), k(X)) = \text{Cov}(f(X_I), g(X_J))$$

□

Weak association also has some nice properties. We list a few of them here.

Proposition 2.1.2. *Let $X = (X_1, ..., X_d)$ be a random vector.*

(a) Weak association is equivalent to inequality (2.6) for f, g non-decreasing indicator functions on $\mathbb{R}^k$, $\mathbb{R}^n$ respectively.

(b) Weak association is preserved under weak convergence, i.e. given a sequence $(X_n)_{n\in\mathbb{N}}$, if X_n is WA for all n, and $X_n \Rightarrow X$, then X is WA.

(c) If X has independent components, then X is WA.

(d) Any subset $X_I = (X_i : i \in I) \subseteq (X_1, ..., X_d)$ is WA if X is WA.

(e) Let $X = (X_1, ..., X_d)$ and $Y = (Y_1, ..., Y_n)$ be independent. If X is WA in $\mathbb{R}^d$ and Y is WA in $\mathbb{R}^n$, then the random vector $(X, Y) := (X_1, ..., X_d, Y_1, ..., Y_n)$ is WA in $\mathbb{R}^{d+n}$.

Proof. The technique of the proofs for these properties is very similar to the technique to show these properties in the case of association. The we omit these proofs and refer the reader to the proofs in Proposition 2.1.1. □

Remark 2.1.2. The previous claims in this section indicate the following implications between association, weak association, and positive correlation:

$$\text{association} \implies \text{weak association} \implies \text{positive correlation.}$$

In some instances, the above three notions may all be equivalent. For example, Esary et. al. showed that for bivariate random vectors (S, T) which have Bernoulli marginals S and T, association of (S, T) is equivalent to positive correlation [16]:

Theorem 2.1 (Esary et. al. [16]). *Let $S \sim Bern(p)$ and $T \sim Bern(q)$, i.e. $\mathbb{P}(S=1)=p$, $\mathbb{P}(S=0)=1-p$ and $\mathbb{P}(T=1)=q$, $\mathbb{P}(T=0)=1-q$. Then (S,T) associated if and only if (S,T) is positively correlated.*

Proof. See Esary et. al. [16]. □

We can also to extend Esary's result to more general random variables that take on two states in some partially ordered state space. In this instance, we can show WA is equivalent to PC.

Theorem 2.2 (T. 2017). *Let S and T be random variables taking on two values $\{a,b\}$, where $a \leq b$ are values on a partially ordered space E. Then X is WA if and only if X is positively correlated.*

Proof. ($\Rightarrow$) True in general.

($\Leftarrow$) Let $f, g : E \to \mathbb{R}$ non-decreasing. Let the joint distributions of (S,T) be given by

	$T=a$	$T=b$
$S=a$	a_{11}	a_{12}
$S=b$	a_{21}	a_{22}

where $a_{11}+a_{12}=p$, $a_{21}+a_{22}=1-p$, $a_{11}+a_{21}=q$, $a_{12}+a_{22}=1-q$. Observe first that

$$
\begin{aligned}
&\mathrm{Cov}(S,T)\\
&= \mathbb{E}(ST) - \mathbb{E}S\mathbb{E}T\\
&= a_{11}a^2 + a_{12}ab + a_{21}ab + a_{22}b^2 - [(a_{11}+a_{12})a + (a_{21}+a_{22})b][(a_{11}+a_{21})a + (a_{12}+a_{22})b]\\
&= a_{11}a^2 + a_{12}ab + a_{21}ab + a_{22}b^2\\
&\quad - (a_{11}+a_{12})(a_{11}+a_{21})a^2 - (a_{21}+a_{22})(a_{11}+a_{21})ab\\
&\quad - (a_{11}+a_{12})(a_{12}+a_{22})ab - (a_{21}+a_{22})(a_{12}+a_{22})b^2\\
&= [a_{11} - (a_{11}+a_{12})(a_{11}+a_{21})]a^2 + [a_{12} - (a_{11}+a_{12})(a_{12}+a_{22})]ab\\
&\quad + [a_{21} - (a_{21}+a_{22})(a_{11}+a_{21})]ab + [a_{22} - (a_{21}+a_{22})(a_{12}+a_{22})]b^2\\
&= Ka^2 - Kab - Kab + Kb^2\\
&= K(b-a)^2
\end{aligned}
$$

where K is the coefficient in front of a^2 and b^2 terms:

$$K := a_{11} - (a_{11} + a_{12})(a_{11} + a_{21}) \overset{(*)}{=} a_{22} - (a_{21} + a_{22})(a_{12} + a_{22})$$

and $-K$ is the coefficient in front of ab terms:

$$-K := a_{12} - (a_{11} + a_{12})(a_{12} + a_{22}) = a_{21} - (a_{21} + a_{22})(a_{11} + a_{21}).$$

It takes some algebra to show the above relations. We just show $(*)$.

$$\begin{aligned}
&a_{11} - (a_{11} + a_{12})(a_{11} + a_{21}) = a_{22} - (a_{21} + a_{22})(a_{12} + a_{22}) \\
\Leftrightarrow\ &a_{11} - pq = a_{22} - (1-p)(1-q) \\
\Leftrightarrow\ &a_{11} - pq = a_{22} - 1 + p + q - pq \\
\Leftrightarrow\ &a_{11} = a_{22} - 1 + p + q \\
\Leftrightarrow\ &a_{11} = a_{22} - 1 + a_{11} + a_{12} + a_{11} + a_{21} \\
\Leftrightarrow\ &a_{11} = a_{11}.
\end{aligned}$$

Thus, if we assume $\mathrm{Cov}(S, T) \geq 0$, then $K \geq 0$. Let's assume for non-triviality that $K > 0$.

$$\begin{aligned}
\mathrm{Cov}(f(S), g(T)) &= a_{11} f(a) g(a) + a_{12} f(a) g(b) + a_{21} f(b) g(a) + a_{22} f(b) g(b) \\
&\quad - [(a_{11} + a_{12}) f(a) + (a_{21} + a_{22}) f(b)][(a_{11} + a_{21}) g(a) + (a_{12} + a_{22}) g(b)] \\
&= a_{11} f(a) g(a) + a_{12} f(a) g(b) + a_{21} f(b) g(a) + a_{22} f(b) g(b) \\
&\quad - (a_{11} + a_{12})(a_{11} + a_{21}) f(a) g(a) - (a_{21} + a_{22})(a_{11} + a_{21}) f(b) g(a) \\
&\quad - (a_{11} + a_{12})(a_{12} + a_{22}) f(a) g(b) - (a_{21} + a_{22})(a_{12} + a_{22}) f(b) g(b) \\
&= [a_{11} - (a_{11} + a_{12})(a_{11} + a_{21})] f(a) g(a) + [a_{12} - (a_{11} + a_{12})(a_{12} + a_{22})] f(a) g(b) \\
&\quad + [a_{21} - (a_{21} + a_{22})(a_{11} + a_{21})] f(b) g(a) + [a_{22} - (a_{21} + a_{22})(a_{12} + a_{22})] f(b) g(b) \\
&= K f(a) g(a) - K f(a) g(b) - K f(b) g(a) + K f(b) g(b) \\
&= K(f(b) - f(a))(g(b) - g(a)) \\
&\geq 0
\end{aligned}$$

since $K > 0$ and f, g non-decreasing. $\square$

In general, the three notions A, WA, and PC are not equivalent. We present the following counterexamples.

Example 2.1.1 (PC but not WA)**.** Let $X = (X_1, X_2, X_3)$, where $X_i \sim Bern(p_i)$. Let the joint distribution of X be given by the following probabilities:

$$p_{ijk} := \mathbb{P}(X_1 = i, X_2 = j, X_3 = k)$$

$$\begin{aligned} p_{100} = p_{010} = p_{001} &= 1/12, \\ p_{110} = p_{101} = p_{011} &= 5/36, \\ p_{000} = 1/3, \quad p_{111} &= 0. \end{aligned}$$

Then X is (PC):

$$\begin{aligned} \mathrm{Cov}(X_1, X_2) &= \mathbb{E}X_1X_2 - \mathbb{E}X_1\mathbb{E}X_2 \\ &= \mathbb{P}(X_1 = 1, X_2 = 1) - \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) \\ &= p_{110} + p_{111} - (p_{100} + p_{110} + p_{101} + p_{111})(p_{010} + p_{110} + p_{011} + p_{111}) \\ &= 5/36 - (1/12 + 5/36 + 5/36)(1/12 + 5/36 + 5/36) \\ &= 11/1296 \end{aligned}$$

Similarly, $\mathrm{Cov}(X_1, X_3) = \mathrm{Cov}(X_2, X_3) = 11/1296$. Now let $f(x_1, x_2) = x_1x_2$ and $g(x_3) = x_3$, which are monotone on $\{0, 1\}^2$ and $\{0, 1\}$, respectively. Partition X into (X_1, X_2) and X_3. Then

$$\begin{aligned} \mathrm{Cov}(f(X_1, X_2), g(X_3)) &= \mathbb{E}f(X_1, X_2)g(X_3) - \mathbb{E}f(X_1, X_2)\mathbb{E}g(X_3) \\ &= \mathbb{E}X_1X_2X_3 - (\mathbb{E}X_1X_2)(\mathbb{E}X_3) \\ &= 0 - (5/36)(1/12 + 5/36 + 5/36) \\ &= -65/1296 \end{aligned}$$

Thus, X is not (WA).

We also present an example that was stated in [16], but was computed clearly in [9].

Example 2.1.2 (WA but not A)**.** Let $X = (X_1, X_2)$, where $X_1, X_2 \in \{0, 1, 2\}$ with the following joint probabilities.

	$X_1 = 0$	$X_1 = 1$	$X_1 = 2$
$X_2 = 0$	15/64	0	8/64
$X_2 = 1$	0	18/64	0
$X_2 = 2$	8/64	0	15/64

X is (WA), i.e. $\mathrm{Cov}(f(X_1), g(X_2)) \geq 0$ for all $f, g : \mathbb{R} \to \mathbb{R}$ non-decreasing. Is is sufficient to check this inequality for non-decreasing indicator functions, i.e. $\mathbb{1}_{[x,\infty)}$, $\mathbb{1}_{[y,\infty)}$, where $x, y \geq 0$. Observe however that for either $x = 0$ or $y = 0$, the inequality becomes trivial. Thus, we need only to check it for when $x, y \in \{1, 2\}$. For example, we compute covariance for $x, y = 1$.

$$\begin{aligned}\mathrm{Cov}(\mathbb{1}_{[1,\infty)}(X_1), \mathbb{1}_{[1,\infty)}(X_2)) &= \mathbb{P}(X_1 \geq 1, X_2 \geq 1) - \mathbb{P}(X_1 \geq 1)\mathbb{P}(X_2 \geq 1) \\ &= 431/4096\end{aligned}$$

We let the reader verify the rest of their own. Now choose $f(x_1, x_2) = \mathbb{1}_{(1,\infty)}(x_1 \vee x_2)$ and $g(x_1, x_2) = \mathbb{1}_{(0,\infty)}(x_1 \wedge x_2)$. They are both non-decreasing functions on $\mathbb{R}^2$, and

$$\begin{aligned}\mathrm{Cov}(f(X_1, X_2), g(X_1, X_2)) &= \mathbb{E}f(X_1, X_2)g(X_1, X_2) - \mathbb{E}f(X_1, X_2)\mathbb{E}g(X_1, X_2) \\ &= 15/64 - (31/64)(33/64) \\ &= -63/4096\end{aligned}$$

Thus, X is not associated.

2.2 Positive supermodular association, supermodular dependence, orthant dependence

Along with association and weak association, there are other forms of positive dependence stronger than PC that are defined based on stochastic ordering. In this section, we define general stochastic orders. Then we discuss how these orders can be used to define various forms of dependence in multivariate distributions, such as positive supermodular dependence (PSD) and positive orthant dependence (POD). We also discuss another related form of positive dependence called positive supermodular association (PSA), that is connected to association and POD. For more on stochastic orders, we refer the reader to [34].

2.2.1 Stochastic orders

Stochastic orders for between multivariate distributions on state space E are usually defined on a space of functions $\mathcal{F}$ contained in a Banach space of functions $\mathbb{B}(E)$ on E. Here, we will just choose $E = \mathbb{R}^d$, and $\mathbb{B}(E) = (B_b(\mathbb{R}^d), ||\cdot||_\infty)$, the bounded functions on $\mathbb{R}^d$ with the sup-norm. Let $M(\mathbb{R}^d)$ be the set of probability measure on $\mathbb{R}^d$. Then we can define a **stochastic order** $\leq_{\mathcal{F}}$ between probability measures:

$$\mu \leq_{\mathcal{F}} \nu \qquad \text{if} \qquad \int_{\mathbb{R}^d} f \, d\mu \leq \int_{\mathbb{R}^d} f \, d\nu$$

for all $f \in \mathcal{F}$ such that the above integrals exist. One can also write this in terms of random vectors X and Y on $\mathbb{R}^d$:

$$X \leq_{\mathcal{F}} Y \qquad \text{if} \qquad \mathbb{E}f(X) \leq \mathbb{E}f(Y)$$

for all $f \in \mathcal{F}$ such that the above expectations exist. Examples of $\mathcal{F}$ include

$$\begin{aligned}
\mathcal{F}_i &= \{f : \mathbb{R}^d \to \mathbb{R}, f \text{ non-decreasing}\} \\
\mathcal{F}_{sm} &= \{f : \mathbb{R}^d \to \mathbb{R}, f \text{ supermodular}\} \\
\mathcal{F}_{ipr} &= \{f : \mathbb{R}^d \to \mathbb{R}, f = \prod_{i=1}^{d} f_i, \ f_i \text{ non-decreasing, positive}\} \\
\mathcal{F}_{dpr} &= \{f : \mathbb{R}^d \to \mathbb{R}, f = \prod_{i=1}^{d} f_i, \ f_i \text{ non-increasing, positive}\}
\end{aligned}$$

$$\mathcal{F}_{uo} = \{f : \mathbb{R}^d \to \mathbb{R}, f(x) = \mathbb{1}_{\{x>t\}},\ t \in \mathbb{R}^d\}$$
$$\mathcal{F}_{lo} = \{f : \mathbb{R}^d \to \mathbb{R}, f(x) = -\mathbb{1}_{\{x\leq t\}},\ t \in \mathbb{R}^d\}$$
$$\mathcal{F}_{ism} = \mathcal{F}_i \cap \mathcal{F}_{sm}.$$

We refer the reader to [34] and [41] for additional examples. A function $f : \mathbb{R}^d \to \mathbb{R}$ is **supermodular** if it satisfies

$$f(x \wedge y) + f(x \vee y) \geq f(x) + f(y), \tag{2.7}$$

where $x \wedge y$ and $x \vee y$ are the componentwise minimum and componentwise maximum, respectively. If $f \in C^2(\mathbb{R}^d)$, then f is supermodular if and only if

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0, \quad \forall i \neq j. \tag{2.8}$$

The spaces $\mathcal{F}_{uo}$ and $\mathcal{F}_{lo}$ are referred to as the **upper orthant** and **lower orthant** orders. One can also define **orthant orders** based on the joint probability distribution function. For a random vector $X = (X_1, ..., X_d)$, define, for $t = (t_1, ..., t_d)$,

$$F_X(t) = \mathbb{P}(X_1 \leq t_1, ..., X_d \leq t_d), \qquad \overline{F}_X(t) = \mathbb{P}(X_1 > t_1, ..., X_d > t_d).$$

Then we say X is less than Y with respect to **upper orthant order** ($X \leq_{uo} Y$) if

$$\overline{F}_X(t) \leq \overline{F}_Y(t), \quad \text{for all } t = (t_1, ..., t_d) \in \mathbb{R}^d.$$

X is less than Y with respect to **lower orthant order** ($X \leq_{lo} Y$) if

$$F_X(t) \leq F_Y(t), \quad \text{for all } t = (t_1, ..., t_d) \in \mathbb{R}^d.$$

It is obvious to see that $X \leq_{\mathcal{F}_{uo}} Y \iff X \leq_{uo} Y$, and similarly for $\leq_{\mathcal{F}_{lo}}$ and $\leq_{lo}$. We will just normally use the subscript in $\mathcal{F}$ to refer to its stochastic order. For example, for **supermodular order**, we will just write $\leq_{sm}$ in place of $\leq_{\mathcal{F}_{sm}}$. Combining the lower and upper orthant orders, we get concordance order: X is less than Y with respect to **concordance order** ($X \leq_c Y$) if

$$X \leq_{uo} Y \text{ and } X \leq_{lo} Y.$$

It can be shown that concordance order $\leq_c$ can be generated by the two spaces $\mathcal{F}_{ipr}$ and $\mathcal{F}_{dpr}$. In other words, $X \leq_c Y$ if and only if $X \leq_{ipr}$ and $X \leq_{dpr} Y$ if and only if $X \leq_{uo} Y$ and $X \leq_{lo} Y$. See references in the proof of Proposition 2.2.1.

2.2.2 Dependence induced from stochastic orders

For defining different forms of positive dependence, we will focus on the supermodular, upper orthant, lower orthant, and concordance orders.

Let $X = (X_1, ..., X_d)$ be a random vector in $\mathbb{R}^d$, and $X^\perp = (X_1^\perp, ..., X_d^\perp)$ a random vector with independent components $X_i^\perp$, and $X_i^\perp \overset{d}{=} X_i$ for all $i \in \{1, ..., d\}$. Definitions 2.3, 2.4, and 2.5 were first proposed by Lehmann [31].

Definition 2.3. X is **positive upper orthant dependent (PUOD)** if $X^\perp \leq_{uo} X$. Equivalently, X satisfies, for all $t = (t_1, ..., t_d) \in \mathbb{R}^d$,

$$\mathbb{P}(X_1 > t_1, ..., X_d > t_d) \geq \mathbb{P}(X_1 > t_1)...\mathbb{P}(X_d > t_d). \tag{2.9}$$

Definition 2.4. X is **positive lower orthant dependent (PLOD)** if $X^\perp \leq_{lo} X$. Equivalently, X satisfies, for all $t = (t_1, ..., t_d) \in \mathbb{R}^d$,

$$\mathbb{P}(X_1 \leq t_1, ..., X_d \leq t_d) \geq \mathbb{P}(X_1 \leq t_1)...\mathbb{P}(X_d \leq t_d). \tag{2.10}$$

Definition 2.5. X is **positive orthant dependent (POD)** if $X^\perp \leq_c X$, i.e. if X is PUOD and PLOD, or equivalently, satisfying (2.9) and (2.10).

Definition 2.6. X is **positive supermodular dependent (PSD)** if $X^\perp \leq_{sm} X$.

For PUOD, PLOD, and POD, we can restate the definition using the space of functions $\mathcal{F}_{ipr}$ and $\mathcal{F}_{dpr}$. This was first proven by Bergmann [5].

Proposition 2.2.1.

(a) X is PUOD if and only if $X^\perp \leq_{ipr} X$.

(b) X is PLOD if and only if $X^\perp \leq_{dpr} X$.

(c) X is POD if and only if $X^\perp \leq_{ipr} X$ and $X^\perp \leq_{dpr} X$.

Proof. For a proof, see [5] or [34]. □

Remark 2.2.1. In conclusion from the above proposition, we can say that X is POD if

$$\mathbb{E}f_1(X_1)...f_d(X_d) \geq \mathbb{E}f_1(X_1)...\mathbb{E}f_d(X_d) \tag{2.11}$$

for all such functions $f_i : \mathbb{R} \to \mathbb{R}_+$ all non-decreasing or all non-increasing.

Here is a nice property of POD originally stated in [34] in terms of stochastic order.

Proposition 2.2.2. *Let $X = (X_1, ..., X_d)$ be POD, and $\Phi = (\Phi_1, ..., \Phi_k) : \mathbb{R}^d \to \mathbb{R}^k$ be defined as $\Phi(x_1, ..., x_d) = (\Phi_1(x_1), ..., \Phi_k(x_k))$, where $k \leq d$, and Φ_i's non-decreasing on $\mathbb{R}$. Then $\Phi(X)$ is POD.*

Proof. Choose $f = \prod_{i=1}^k f_i$, where $f_i : \mathbb{R} \to \mathbb{R}_+$ all non-decreasing. According to Proposition 2.2.1 we want to show that

$$\mathbb{E}\prod_{i=1}^k f_i(\Phi_i(X_i)) \geq \prod_{i=1}^k \mathbb{E}f_i(\Phi_i(X_i)).$$

For all $i \leq k$, we have $f_i \circ \Phi_i : \mathbb{R} \to \mathbb{R}$ is also non-decreasing because Φ is non-decreasing. Thus $g : \mathbb{R}^d \to \mathbb{R}$ defined by

$$g(x) := f \circ \Phi(x) = f \circ \Phi(x_1, ..., x_d) = f(\Phi_1(x_1), ..., \Phi_k(x_k)) = \prod_{i=1}^k f_i(\Phi_i(x_i))$$

is a product-type function $g = \prod_{i=1}^d g_i$ with non-decreasing components $g_i := f_i \circ \Phi_i$. Thus, by POD of X, we have

$$\begin{aligned}\mathbb{E}\prod_{i=1}^k f_i(\Phi_i(X_i)) &= \mathbb{E}\prod_{i=1}^k g_i(X_i) \\ &\geq \prod_{i=1}^k \mathbb{E}g_i(X_i)\end{aligned}$$

$$= \prod_{i=1}^{k} \mathbb{E} f_i(\Phi_i(X_i)).$$

A similar result holds if we make all f_i's non-increasing. □

Remark 2.2.2. (i) Observe that we can generalize the above theorem to be that $\Phi : \mathbb{R}^d \to \mathbb{R}^k$ is a monotone transformation whose components Φ_i only act on a single component X_i of the random vector X. So for example, Φ could look like $\Phi(X) = (\Phi_2(X_2), \Phi_3(X_3), \Phi_7(X_7))$.

(ii) Proposition 2.2.2 still holds if we replace POD with PUOD and PLOD.

One also obtains a similar result when we look at certain monotone transformations of PSD random vectors.

Lemma 2.2.1. *If $X = (X_1, ..., X_d)$ and $Y = (Y_1, ..., Y_d)$ are random vectors with independent components, and $X_i \overset{d}{=} Y_i$ for all i, then*

$$\mathbb{E} f(X_1, ..., X_d) = \mathbb{E} f(Y_1, ..., Y_d)$$

for all f such that the expectations exist. Moreover, $X \overset{d}{=} Y$.

Proposition 2.2.3. *Let $X = (X_1, ..., X_d)$ be PSD. Then for any $g_1, ..., g_d : \mathbb{R} \to \mathbb{R}$ non-decreasing, $(g_1(X_1), ..., g_d(X_d))$ is also PSD.*

Proof. Let $f \in \mathcal{F}_{ism}$. We want to show

$$\mathbb{E} f(g_1(X_1), ..., g_d(X_d)) \geq \mathbb{E} f(g_1(X_1)^{\perp}, ..., g_d(X_d)^{\perp}) \tag{2.12}$$

where $(g_1(X_1)^{\perp}, ..., g_d(X_d)^{\perp})$ has marginals $g_i(X_i)^{\perp} \overset{d}{=} g_i(X_i)$, and independent components. Observe that since $X^{\perp}$ has independent components, then $g_i(X_i^{\perp})$ will be independent for all $i \in \{1, ..., d\}$, and $g_i(X_i^{\perp}) \overset{d}{=} g_i(X_i)$ for all i. Then we can just consider, for the RHS of (2.12), the vector $(g_1(X_1^{\perp}), ..., g_d(X_d^{\perp}))$, since the independent components will yield $\mathbb{E} f(g_1(X_1)^{\perp}, ..., g_d(X_d)^{\perp}) = \mathbb{E} f(g_1(X_1^{\perp}), ..., g_d(X_d^{\perp}))$ by Lemma 2.2.1. Then (2.12) becomes

$$\mathbb{E}(f \circ g)(X) = \mathbb{E} f(g_1(X_1), ..., g_d(X_d)) \geq \mathbb{E} f(g_1(X_1^{\perp}), ..., g_d(X_d^{\perp})) = \mathbb{E}(f \circ g)(X^{\perp}) \tag{2.13}$$

where $g : \mathbb{R}^d \to \mathbb{R}^d$ is defined by $g(x_1, ..., x_d) = (g_1(x_1), ..., g_d(x_d))$. Thus it suffices to show $f \circ g$ is supermodular. We prove this for $d = 2$, but this can be easily generalized to higher dimensions. WLOG, consider $x = (x_1, x_2)$, $y = (y_1, y_2)$, where $x_1 < y_1$ and $x_2 > y_2$. Then this implies $g_1(x_1) \leq g_1(y_1)$ and $g_2(x_2) \geq g_2(y_2)$. Hence,

$$
\begin{aligned}
& f(g(x \wedge y)) + f(g(x \vee y)) - f(g(x)) - f(g(y)) \\
&= f(g_1(x_1 \wedge y_1), g_2(x_2 \wedge y_2)) + f(g_1(x_1 \vee y_1), g_2(x_2 \vee y_2)) \\
&\quad - f(g_1(x_1), g_2(x_2)) - f(g_1(y_1), g_2(y_2)) \\
&= f(g_1(x_1), g_2(y_2)) + f(g_1(y_1), g_2(x_2)) \\
&\quad - f(g_1(x_1), g_2(x_2)) - f(g_1(y_1), g_2(y_2)) \\
&= f(g_1(x_1) \wedge g_1(y_1), g_2(x_2) \wedge g_2(y_2)) + f(g_1(x_1) \vee g_1(y_1), g_2(x_2) \vee g_2(y_2)) \\
&\quad - f(g_1(x_1), g_2(x_2)) - f(g_1(y_1), g_2(y_2)) \\
&\geq 0
\end{aligned}
$$

where we obtain the last equality by the supermodularity of f on the pairs $(g_1(x_1), g_2(x_2))$ and $(g_1(y_1), g_2(y_2))$. So function $f \circ g$ is supermodular, and thus by the PSD of X, we get our desired inequality (2.13). □

The space of functions $\mathcal{F}$, which help to define a particular stochastic order $\leq_{\mathcal{F}}$, can yield interesting forms of positive dependence, like PUOD, PLOD, POD, and PSD given above. Interesting uses and applications of these forms of positive dependence can be found in [34], [4]. Additionally, these some of these spaces of functions can also determine their own form of positive dependence, but not through the means of a stochastic order. Two examples are association and positive supermodular association (PSA).

Association can be defined by the functions $\mathcal{F}_i$, i.e. X associated if

$$\text{Cov}(f(X), g(X)) \geq 0, \quad f, g \in \mathcal{F}_i$$

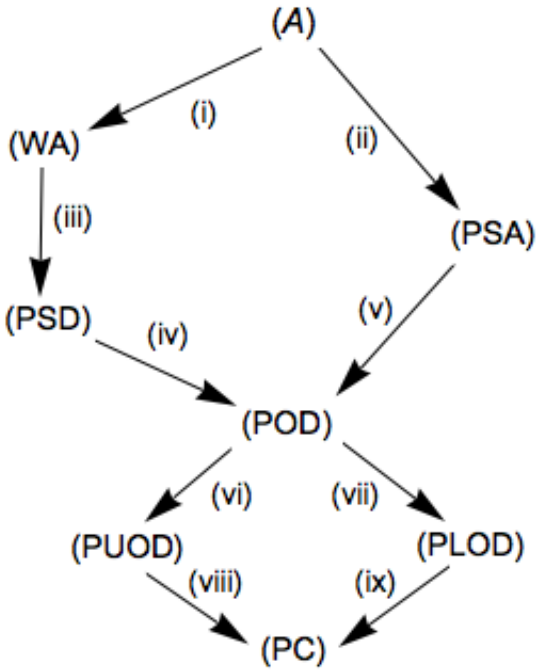

Figure 2.1: Implication map of positive dependence

such that the covariance is defined. In a similar way, one can also define positive supermodular association.

Definition 2.7. $X = (X_1, ..., X_d)$ is **positive supermodular associated (PSA)** if

$$\text{Cov}(f(X), g(X)) \geq 0, \quad f, g \in \mathcal{F}_{ism} \tag{2.14}$$

such that the covariance is defined

For applications of PSA, we refer the reader to [42]. Note that by Remark 2.1.2, we have relations between association, weak association, and positive correlation. Want to also show relations between PSA, PSD, POD, PUOD, PLOD, and where they fit in with association, WA, and PC. The implication map in Figure 2.1 will help us organize the different strengths of the different forms of dependence. Our intent is to show each of the implications, reveal examples where certain notions may be equivalent, and provide some counter-examples to show that they, in general, are not equivalent.

We've shown in the previous subsection the implications given by (i) and WA $\Rightarrow$ PC, and examples. We would now like to show the rest of the implications given in Figure 2.1. We first require the following lemmas.

Lemma 2.2.2. *Define $f : \mathbb{R}^d \to \mathbb{R}$ by $f(x_1, ..., x_d) := f_1(x_1)...f_d(x_d)$, where f_i are non-decreasing and non-negative. Then f is supermodular. (The same result holds if we replace "non-decreasing" by "non-increasing").*

Proof. We just show the proof for $d = 2$. Let $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$ such that, WLOG, $x_1 < y_1$ and $x_2 > y_2$. Then

$$\begin{aligned} f(x \wedge y) + f(x \vee y) &= f(x_1 \wedge y_1, x_2 \wedge y_2) + f(x_1 \vee y_1, x_2 \vee y_2) \\ &= f_1(x_1 \wedge y_1)f_2(x_2 \wedge y_2) + f_1(x_1 \vee y_1)f_2(x_2 \vee y_2) \\ &= f_1(x_1)f_2(y_2) + f_1(y_1)f_2(x_2). \end{aligned}$$

Thus,

$$\begin{aligned} f(x \wedge y) + f(x \vee y) - f(x) - f(y) &= f_1(x_1)f_2(y_2) + f_1(y_1)f_2(x_2) - f_1(x_1)f_2(x_2) - f_1(y_1)f_2(y_2) \\ &= f_1(x_1)[f_2(y_2) - f_2(x_2)] + f_1(y_1)[f_2(x_2) - f_2(y_2)] \\ &= f_1(x_1)[f_2(y_2) - f_2(x_2)] - f_1(y_1)[f_2(y_2) - f_2(s_2)] \\ &= [f_2(y_2) - f_2(x_2)][f_1(x_1) - f_1(y_1)] \\ &\geq 0 \end{aligned}$$

by monotonicity of f_1 and f_2. □

Lemma 2.2.3. *Let $X = (X_1, ..., X_d)$ be POD (PUOD, PLOD, respectively). For all $\{s_1, ..., s_n\} \subseteq \{1, ..., d\}$, we have that $(X_{s_1}, ..., X_{s_n})$ is POD (PUOD, PLOD, respectively).*

Proof. Let $(f_j)_{j=1}^n$ be non-decreasing functions, where $f_j : \mathbb{R} \to \mathbb{R}_+$. For all $i \in \{1, ..., d\} \setminus \{s_1, ..., s_n\}$, set $f_i = 1$. Then these f_i are non-decreasing, non-negative. With $f = \prod_{i=1}^d f_i$, by POD, we have

$$\prod_{j=1}^n \mathbb{E} f_j(X_{s_j}) = \prod_{j=1}^n \mathbb{E} f_j(X_{s_j}^\perp) = \mathbb{E}\left[\prod_{j=1}^n f_j(X_{s_j}^\perp)\right] = \mathbb{E} f(X^\perp)$$

$$\leq \mathbb{E}f(X)$$
$$= \mathbb{E}\left[\prod_{j=1}^{n} f_j(X_{s_j})\right].$$

Thus $(X_{s_1}, ..., X_{s_n})$ is PUOD by Proposition 2.2.1 (a). Now to show PLOD, we can replace f_j "non-decreasing" by "non-increasing. This yields $(X_{s_1}, ..., X_{s_n})$ is POD. □

Proposition 2.2.4.

(i) (A) implies (WA)

(ii) (A) implies (PSA)

(iii) (WA) implies (PSD)

(iv) (PSD) implies (POD)

(v) (PSA) implies (POD)

(vi) (POD) implies (PUOD)

(vii) (POD) implies (PLOD)

(viii) (PUOD) implies (PC)

(ix) (PLOD) implies (PC)

Proof. Let $X = (X_1, ..., X_d)$ and $X^{\perp} = (X_1^{\perp}, ..., X_d^{\perp})$, where components of $X^{\perp}$ are independent, and $X_i^{\perp} \overset{d}{=} X_i$ for all $i \in \{1, ..., d\}$.

(i) We proved this in the previous section.

(ii) Let $f, g \in \mathcal{F}_{ism} \subset \mathcal{F}_i$. This implies $\text{Cov}(f(X), g(X)) \geq 0$ by association.

(iii) The proof for this is quite tedious, and we refer the reader to the paper by Christofides and Vaggelatou [12].

(iv) We know, by PSD, f supermodular implies $\mathbb{E}f(X^{\perp}) \leq \mathbb{E}f(X)$. Fix $t = (t_1, ..., t_d)$. Define $f(x_1, ..., x_d) := \mathbb{1}_{(t_1,\infty)}(x_1)...\mathbb{1}_{(t_d,\infty)}(x_d)$. This function is supermodular by Lemma 2.2.2, since $\mathbb{1}_{(t_i,\infty)}(x_i)$ is non-decreasing and non-negative for all i. Hence, by PSD,

$$\begin{aligned}\mathbb{P}(X_1 > t_1)...\mathbb{P}(X_d > t_d) &= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1)]...\mathbb{E}[\mathbb{1}_{(t_d,\infty)}(X_d)] \\ &= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1^{\perp})]...\mathbb{E}[\mathbb{1}_{(t_d,\infty)}(X_d^{\perp})]\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1^{\perp})...\mathbb{1}_{(t_d,\infty)}(X_d^{\perp})], \text{ by independence} \\
&= \mathbb{E}[f(X^{\perp})] \\
&\leq \mathbb{E}[f(X)] \\
&= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1)...\mathbb{1}_{(t_d,\infty)}(X_d)] \\
&= \mathbb{P}(X_1 > t_1, ..., X_d > t_d),
\end{aligned}
$$

giving us our desired result. We can do a similar trick to show the lower orthant dependence by considering $f_i = \mathbb{1}_{(-\infty,t_i]}$, which are non-increasing, and using Lemma 2.2.2.

(v) We prove this by induction. Base Case: $d = 2$. We know $\mathbb{E}f(X)g(X) \geq \mathbb{E}f(X)\mathbb{E}g(X)$ for all $f, g \in \mathcal{F}_{ism}$. Fix $t = (t_1, t_2) \in \mathbb{R}^2$. Define $f(x_1, x_2) = \mathbb{1}_{(t_1,\infty)}(x_1)$ and $g(x_1, x_2) = \mathbb{1}_{(t_2,\infty)}(x_2)$. Then both f, g are non-decreasing, supermodular. Hence,

$$
\begin{aligned}
\mathbb{P}(X_1 > t_1, X_2 > t_2) &= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1)\mathbb{1}_{(t_2,\infty)}(X_2)] \\
&= \mathbb{E}[f(X)g(X)] \\
&\geq \mathbb{E}f(X)\mathbb{E}g(X) \\
&= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1)]\mathbb{E}[\mathbb{1}_{(t_2,\infty)}(X_2)] \\
&= \mathbb{P}(X_1 > t_1)\mathbb{P}(X_2 > t_2)
\end{aligned}
$$

Induction Hypothesis: Assume the desired inequality holds for $d-1$, i.e. $\mathbb{P}(X_1 > t_1, ..., X_{d-1} > t_{d-1}) \geq \mathbb{P}(X_1 > t_1)...\mathbb{P}(X_{d-1} > t_{d-1})$. Inductive Step: Choose $f(x_1, ..., x_d) = \mathbb{1}_{(t_1,\infty)}(x_1)...\mathbb{1}_{(t_{d-1},\infty)}(x_{d-1})$, and $g(x_1, ..., x_d) = \mathbb{1}_{(t_d,\infty)}(x_d)$. Then by Lemma 2.2.2, $f, g \in \mathcal{F}_{ism}$. Hence,

$$
\begin{aligned}
\mathbb{P}(X_1 > t_1, ..., X_{d-1} > t_{d-1}, X_d > t_d) &= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1)...\mathbb{1}_{(t_{d-1},\infty)}(X_{d-1})\mathbb{1}_{(t_d,\infty)}(X_d)] \\
&= \mathbb{E}f(X)g(X) \\
&\geq \mathbb{E}f(X)\mathbb{E}g(X) \\
&= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1)...\mathbb{1}_{(t_{d-1},\infty)}(X_{d-1})]\,\mathbb{E}[\mathbb{1}_{(t_d,\infty)}(X_d)]
\end{aligned}
$$

$$= \mathbb{P}(X_1 > t_1, ..., X_{d-1} > t_{d-1})\, \mathbb{P}(X_d > t_d)$$
$$\geq \mathbb{P}(X_1 > t_1)...\mathbb{P}(X_{d-1} > t_{d-1})\mathbb{P}(X_d > t_d),$$

where we obtain the last inequality by Induction Hypothesis. To show the lower orthant dependence, we would define our functions f, g similar to those above, except replacing $\mathbb{1}_{(t_k,\infty)}(x_k)$ with $\mathbb{1}_{(-\infty,t_k]}(x_k)$, which are non-increasing supermodular functions, and then use the fact that PSA also means $\mathbb{E}f(X)g(X) \geq \mathbb{E}f(X)\mathbb{E}g(X)$ for f, g supermodular, non-increasing.

(vi) True by definition of POD.

(vii) True by definition of POD.

(viii) Assume $\text{Cov}(X_i, X_j) < \infty$ for all i, j. If X is PUOD, then every bivariate vector (X_i, X_j) is PUOD by Lemma 2.2.3. Thus,

$$\begin{aligned}\text{Cov}(X_i, X_j) &= \int_{\mathbb{R}}\int_{\mathbb{R}} \text{Cov}(\mathbb{1}_{(t,\infty)}(X_i), \mathbb{1}_{(s,\infty)}(X_j))ds\, dt \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} [\mathbb{P}(X_i > t, X_j > s) - \mathbb{P}(X_i > t)\mathbb{P}(X_j > s)]ds\, dt \\ &\geq 0\end{aligned}$$

since the integrand is greater than or equal to 0 by (X_i, X_j) PUOD.

(ix) Assume $\text{Cov}(X_i, X_j) < \infty$ for all i, j. If X is PLOD, then every bivariate vector (X_i, X_j) is PLOD by Lemma 2.2.3. Additionally, (X_i, X_j) is PLOD if and only if (X_i, X_j) is PUOD, since PLOD and PUOD are equivalent notions in dimension 2 (see [34]). Hence, $\text{Cov}(X_i, X_j) \geq 0$ by the same argument as in part (viii).

□

These various forms of positive dependence are in general not equivalent. There are certain conditions and examples in which they do become equivalent.

Proposition 2.2.5. *Let $d = 2$. Then the following are equivalent for the bivariate vector $X = (X_1, X_2)$:*

(i) X is PLOD

(ii) X is PUOD

(iii) X is POD

(iv) X is PSD

(v) X is WA

Proof. We will just show that (i) implies (ii) and then (ii) implies (v).
$((i) \Rightarrow (ii))$. Assume $\mathbb{P}(X_1 \leq s, X_2 \leq t) \geq \mathbb{P}(X_1 \leq s)\mathbb{P}(X_2 \leq t)$. By inclusion-exclusion principle,

$$\begin{aligned}
\mathbb{P}(X_1 > s, X_2 > t) &= 1 - \mathbb{P}(\{X_1 \leq s\} \cup \{X_2 \leq t\}) \\
&= 1 - (\mathbb{P}(X_1 \leq s) + \mathbb{P}(X_2 \leq t) - \mathbb{P}(X_1 \leq s, X_2 \leq t)) \\
&= 1 - \mathbb{P}(X_1 \leq s) - \mathbb{P}(X_2 \leq t) + \mathbb{P}(X_1 \leq s, X_2 \leq t) \\
&\geq 1 - \mathbb{P}(X_1 \leq s) - \mathbb{P}(X_2 \leq t) + \mathbb{P}(X_1 \leq s)\mathbb{P}(X_2 \leq t) \\
&= (1 - \mathbb{P}(X_1 \leq s))(1 - \mathbb{P}(X_2 \leq t)) \\
&= \mathbb{P}(X_1 > s)\mathbb{P}(X_2 > t)
\end{aligned}$$

$((ii) \Rightarrow (v))$. Assume (X_1, X_2) is PUOD. We want to show $\mathrm{Cov}(f(X_1), g(X_2)) \geq 0$ for all $f, g : \mathbb{R} \to \mathbb{R}$ non-decreasing. If f nondecreasing on $\mathbb{R}$, then observe that for any fixed $s \in \mathbb{R}$, $\Phi_s := \mathbb{1}_{(s,\infty)} \circ f : \mathbb{R} \to \mathbb{R}$ is non-decreasing. Hence, for a given $s, t \in \mathbb{R}$, $\Phi(X_1, X_2) = (\Phi_s(X_1), \Phi_t(X_2)) = (\mathbb{1}_{(s,\infty)}(f(X_1)), \mathbb{1}_{(t,\infty)}(g(X_2))$ is PUOD by Proposition 2.2.2 and is thus also PC. Hence

$$\begin{aligned}
\mathrm{Cov}(f(X_1), g(X_2)) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{Cov}(\mathbb{1}_{(s,\infty)}(f(X_1)), \mathbb{1}_{(t,\infty)}(g(X_2)))ds\, dt \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{Cov}(\Phi_s(X_1), \Phi_t(X_2))ds\, dt \\
&\geq 0,
\end{aligned}$$

because $\mathrm{Cov}(\Phi_s(X_1), \Phi_t(X_2)) \geq 0$ by PC.

Thus, by Proposition 2.2.4, we have the desired equivalences. □

The implications in Figure 2.1 are strict. We provide the following examples to demonstrate why for a few of the implications.

Example 2.2.1 ((PC) but not (PUOD))**.** We can actually take the same example of $X = (X_1, X_2, X_3)$ from Example 2.1.1. This distribution is PC as was shown in that example, but is not PUOD. For example, take $0 < t_1, t_2, t_3 < 1$. Then

$$\begin{aligned}\mathbb{P}(X_1 > t_1, X_2 > t_2, X_3 > t_3) &= \mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1) \\ &= p_{111} = 0\end{aligned}$$

$$\begin{aligned}&\mathbb{P}(X_1 > t_1)\mathbb{P}(X_2 > t_2)\mathbb{P}(X_3 > t_3) \\ &= (p_{100} + p_{110} + p_{101})(p_{010} + p_{110} + p_{011})(p_{001} + p_{101} + p_{011}) \\ &= (1/12 + 5/36 + 5/36)(1/12 + 5/36 + 5/36)(1/12 + 5/36 + 5/36) \\ &= 2197/46656\end{aligned}$$

Thus, X is not PUOD.

Example 2.2.2 ((PUOD) but not (POD), (PLOD))**.** The idea behind this example comes from [34] in the form of stochastic orders. Let $X = (X_1, X_2, X_3)$ and $p_{ijk} = \mathbb{P}(X_1 = i, X_2 = j, X_3 = k)$. The distribution of X is given by

$$p_{100} = p_{010} = p_{001} = p_{111} = 1/4$$

Let $X^{\perp} = (X_1^{\perp}, X_2^{\perp}, X_3^{\perp})$ and $q_{ijk} = \mathbb{P}(X_1^{\perp} = i, X_2^{\perp} = j, X_3^{\perp} = k)$

$$q_{ijk} = 1/8, \quad \forall i, j, k \in \{0, 1\}$$

where $X^{\perp}$ has independent components. Additionally, $X_i \overset{d}{=} X_i^{\perp} \sim Bern(1/2)$ for all i. Then

$$\mathbb{P}(X_1 > t_1, X_2 > t_2, X_3 > t_3) \geq \mathbb{P}(X_1^{\perp} > t_1, X_2^{\perp} > t_2, X_3^{\perp} > t_3)$$

for all $t = (t_1, t_2, t_3) \in \mathbb{R}^3$. But for $t = (1/2, 1/2, 1/2)$, then

$$\begin{aligned}\mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, X_3 \leq t_3) = 0 < 1/8 = \mathbb{P}(X_1^\perp \leq t_1)\mathbb{P}(X_2^\perp \leq t_2)\mathbb{P}(X_3^\perp \leq t_3) \\ = \mathbb{P}(X_1^\perp \leq t_1, X_2^\perp \leq t_2, X_3^\perp \leq t_3)\end{aligned}$$

Thus X is PUOD, but is not PLOD, therefore, also not POD.

Example 2.2.3 ((PSD) but not (POD))**.** The idea behind this example comes from [34] in the form of stochastic orders. Let $X = (X_1, X_2, X_3)$ and $p_{ijk} = \mathbb{P}(X_1 = i, X_2 = j, X_3 = k)$. The distribution of X is given by

$$p_{222} = p_{211} = p_{121} = p_{112} = p_{202} = p_{000} = 1/6.$$

Let $X^\perp = (X_1^\perp, X_2^\perp, X_3^\perp)$ and $q_{ijk} = \mathbb{P}(X_1^\perp = i, X_2^\perp = j, X_3^\perp = k)$, with independent components. Thus the distribution can be determined by the marginals

$$\begin{aligned}&\mathbb{P}(X_1^\perp = 2) = 1/2, \quad \mathbb{P}(X_1^\perp = 1) = 1/3, \quad \mathbb{P}(X_1^\perp = 0) = 1/6 \\ &\mathbb{P}(X_2^\perp = 2) = 1/3, \quad \mathbb{P}(X_2^\perp = 1) = 1/3, \quad \mathbb{P}(X_2^\perp = 0) = 1/3 \\ &\mathbb{P}(X_3^\perp = 2) = 1/2, \quad \mathbb{P}(X_3^\perp = 1) = 1/3, \quad \mathbb{P}(X_3^\perp = 0) = 1/6,\end{aligned}$$

and $X_i^\perp \stackrel{d}{=} X_i$. It can be easily shown that

$$\begin{aligned}\mathbb{P}(X_1 > t_1, X_2 > t_2, X_3 > t_3) \geq \mathbb{P}(X_1^\perp > t_1, X_2^\perp > t_2, X_3^\perp > t_3) \\ \mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, X_3 \leq t_3) \geq \mathbb{P}(X_1^\perp \leq t_1, X_2^\perp \leq t_2, X_3^\perp \leq t_3)\end{aligned}$$

for all $t = (t_1, t_2, t_3) \in \mathbb{R}^3$. Thus X is POD. Now define $f(x_1, x_2, x_3) = \max\{x_1+x_2+x_3-4, 0\}$. $f \in \mathcal{F}_{sm}$, and

$$\mathbb{E}f(X) = \mathbb{E}\max(X_1 + X_2 + X_3, 0) = 2p_{222} + p_{221} + p_{212} + p_{122} = 2p_{222} = 1/3$$

and

$$\begin{aligned}\mathbb{E}f(X^\perp) &= \mathbb{E}\max(X_1^\perp + X_2^\perp + X_3^\perp, 0) \\ &= 2q_{222} + q_{221} + q_{212} + q_{122}\end{aligned}$$

$$= 2(1/12) + 1/18 + 1/12 + 1/18 = 13/36$$

Thus, $\mathbb{E}f(X) < \mathbb{E}f(X^{\perp})$, and therefore X is not PSD.

2.3 Positive dependence and infinite divisibility

The examples we've looked at in this chapter thus far have been discrete random vectors. We want to look at more interesting distributions, in particular, infinitely divisible (ID) distributions. We define infinitely divisible distributions, provide examples, and then show sufficient and/or necessary conditions for certain forms of positive dependence for ID distributions. Our standard references in this section are [2], [47], [24].

2.3.1 Infinitely divisible (ID) distributions

Definition 2.8. A random vector $X = (X_1, ..., X_d)$ in $\mathbb{R}^d$ is **infinitely divisible (ID)** if, for all $n \in \mathbb{N}$, there exist $Y_{1,n}, ..., Y_{n,n}$ independent and identically distributed random vectors such that

$$X \stackrel{d}{=} Y_{1,n} + ... + Y_{n,n}. \tag{2.15}$$

Here are some examples of distributions that are infinitely divisible.

Example 2.3.1. Gaussian distributions. $X \sim N(b, \Sigma)$, where $b \in \mathbb{R}^d$ is the mean vector and Σ is the covariance matrix, which is a symmetric positive semidefinite matrix. To see that it is infinitely divisible, for $n \in \mathbb{N}$, in (2.15), choose $(Y_{i,n})_{i=1}^n \stackrel{iid}{\sim} N(b/n, \Sigma/n)$.

Example 2.3.2. Poisson distributions. $X \sim Poiss(\lambda)$, where $\lambda > 0$. To see that it is infinitely divisible, for $n \in \mathbb{N}$, in (2.15), choose $(Y_{i,n})_{i=1}^n \stackrel{iid}{\sim} Poiss(\lambda/n)$.

Example 2.3.3. Compound Poisson distributions. $X = \sum_{j=1}^N Z_j$, where $Z_j \stackrel{iid}{\sim} \mu$, $N \sim Poiss(\lambda)$, and N and $(Z_j)_{j\in\mathbb{N}}$ are independent. To see that it is infinitely divisible, for $n \in \mathbb{N}$, define independent random vectors $(Y_{i,n})_{i=1}^n$ in (2.15) by $Y_{i,n} = \sum_{j=1}^{N^{(n)}} Z_j$, where $N^{(n)} \sim Poiss(\lambda/n)$ are independent of $(Z_j)_{j\in\mathbb{N}}$ for all n.

The following theorem gives a characterization of infinitely divisible distributions.

Theorem 2.3 (Lévy-Khintchine)**.** *A random vector X on $\mathbb{R}^d$ is infinitely divisible if and only if there exists a unique triplet (b, Σ, ν), where $b \in \mathbb{R}^d$, Σ is a symmetric positive semidefinite $d \times d$ matrix, and ν is a σ-finite measure on $\mathbb{R}^d \setminus \{0\}$ with $\int (1 \wedge |y|^2)\nu(dy) < \infty$, such that the characteristic function looks like*

$$\phi_X(u) = \exp\left(iu \cdot b - \frac{1}{2} u \cdot \Sigma u + \int_{y \neq 0} (e^{iu \cdot y} - 1 - iu \cdot y\chi(y))\nu(dy) \right). \tag{2.16}$$

*where $\chi : \mathbb{R}^d \to \mathbb{R}$ is the cut-off function. We call measure ν the **Lévy measure**.*

Different authors prefer various cut-off functions χ. Some authors want their cut-off functions to be smooth with compact support, i.e. $\chi \in C_c^\infty(\mathbb{R}^d)$. The standard cut-off function seen in most literature however is $\chi(y) = \mathbb{1}_{(0,1)}(|y|)$. Changing cut-off functions will only effect parameter b. See [47] for details. Throughout this dissertation, we will write $\chi(y)$ to represent $\mathbb{1}_{(0,1)}(|y|)$ unless we specify otherwise.

The triplet (b, Σ, ν) is often referred to as the the **characteristic triplet** or **Lévy triplet**. We often write $X \sim ID(b, \Sigma, \nu)$ to say X has those characteristics. To interpret these parameters, observe that if we set $\nu = 0$, then (2.16) becomes the characteristic function of a Gaussian distribution with mean b and covariance Σ. If we set $b, \Sigma = 0$ (in $\mathbb{R}^d$ and $\mathbb{R}^{d \times d}$, respectively), and let ν be a finite measure, then (2.16) becomes the characteristic function of a compound Poisson distribution. For more general ν that is not finite, then (2.16) becomes the characteristic function of a distribution that can be constructed as the limit of compound Poisson distributions, see [2]. Thus, in general, we call b, Σ the parameters describing the **Gaussian** component of X and parameter ν describing the **Poissonian** component of X. Hence, any infinitely divisible distribution X can be decomposed into the independent sum: $X = Y + Z$, where Y has Gaussian distribution and Z has Poissonian distribution (Poissonian means compound Poisson or limit of compound Poissons). For more explicit decomposition, see Theorem 2.6.

The exponent of (2.16) is called the **Lévy exponent** or **Lévy symbol**. We will denote this by ψ, where $\psi : \mathbb{R}^d \to \mathbb{C}$ given by

$$\psi(u) = iu \cdot b - \frac{1}{2} u \cdot \Sigma u + \int_{y \neq 0} (e^{iu \cdot y} - 1 - iu \cdot y\chi(y))\nu(dy). \tag{2.17}$$

The function $-\psi$ is a **continuous negative definite function (cndf)** in the sense of Schoenberg, i.e. for all $n \in \mathbb{N}$, $\forall x_1, ..., x_n \in \mathbb{R}^d$ and $\forall z_1, ..., z_n \in \mathbb{C}$ such that $\sum_{j=1}^n z_j = 0$, we have

$$-\sum_{j,k=1}^{n} z_j \overline{z_k} \psi(x_j - x_k) \leq 0. \tag{2.18}$$

(We have a negative out in front of LHS of (2.18) since we called $-\psi$ cndf.) It can be shown that there is a 1-1 correspondence between continuous negative definite functions and the Lévy-Khintchine formula.

Theorem 2.4. *A function $-\psi : \mathbb{R}^d \to \mathbb{C}$ is continuous negative definite if and only if there exists (b, Σ, ν) as given in Theorem 2.3 such that*

$$\psi(u) = iu \cdot b - \frac{1}{2} u \cdot \Sigma u + \int_{y \neq 0} (e^{iu \cdot y} - 1 - iu \cdot y\chi(y))\nu(dy). \tag{2.19}$$

Proof. See [24]. □

Remark 2.3.1. Continuous negative definite functions actually have a Lévy-Khintchine form like (2.17) but with an additional characteristic $a \in \mathbb{R}_+$. In other words, cndf $-\psi$ will have a corresponding characteristic quadruplet (a, b, Σ, ν), where $a \in \mathbb{R}_+$, and ψ will look like the [RHS of (2.17) $- a$]. We just consider the triplet (b, Σ, ν) in this chapter since just the triplet (b, Σ, ν) yields a characteristic function of a probability distribution, whereas including $a > 0$ would not. The parameter a will become more important in Chapter 3.

2.3.2 Positive dependence in ID distributions

The Lévy triplet (b, Σ, ν) fully describes the distribution of infinitely divisible X. Hence, we can look to these parameters to characterize the positive dependence of ID distributions. The following theorem was proven by Pitt (1982) [38] and gives a characterization of association for Gaussian distributions.

Theorem 2.5 (Pitt, [38])**.** *$X \sim N(b, \Sigma)$ in $\mathbb{R}^d$. Then X associated if and only if X is positively correlated, i.e. $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j) \geq 0$ for all $i, j \in \{1, ..., d\}$.*

The above theorem characterizes association for ID distributions of the type $(b, \Sigma, 0)$. What is remarkable about this result is that, for Gaussian distributions, the strongest form of positive dependence that we've discussed, association, is equivalent to the weakest form of positive dependence, positive correlation. We cannot say the same for ID distributions of Poissonian type $(0, 0, \nu)$. Firstly, how can we study association in ID Poissonian distributions $(0, 0, \nu)$? We present a sufficient condition for association of Poissonian distributions, as given by Resnick [39]. We present our own proof, but first require an important theorem about infinitely divisible distributions.

Theorem 2.6 (Lévy-Itô decomposition)**.** *X is an infinitely divisible random vector if and only if there exist a vector $b \in \mathbb{R}^d$, a Gaussian random vector $Y_\Sigma \sim N(0, \Sigma)$, and a Poisson random measure N on $\mathbb{R}^d \setminus \{0\}$ with intensity measure ν, ν being a Lévy measure on $\mathbb{R}^d \setminus \{0\}$, such that*

$$X = b + Y_\Sigma + \int_{|y|<1} y\tilde{N}(dy) + \int_{|y|\geq 1} yN(dy), \tag{2.20}$$

where $\tilde{N}$ is the compensated Poisson random measure, defined by $\tilde{N}(A) := N(A) - \nu(A)$, for $A \in \mathcal{B}(\mathbb{R}^d)$. The parameters (b, Σ, ν) are the same as the ones corresponding to the Lévy triplet in (2.16).

Proof. See Applebaum [2]. □

The Poisson integrals on the RHS of 2.20 are representations of Poissonian distributions. The first one is a limit of compound Poisson distributions, and the second one is a compound Poisson distribution. Again, see [2]. **Poisson random measures (PRM)** N with intensity measure ν on $(\mathbb{R}^d \setminus \{0\}, \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ are Poisson point processes (PPP) on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that satisfy

- $N : \mathcal{B}(\mathbb{R}^d \setminus \{0\}) \times \Omega \to \mathbb{R}_+$,
- $N(A) \sim Poiss(\nu(A))$ if $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ and $\nu(A) < \infty$,
- $\forall A_1, ..., A_n \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$, A_i disjoint, we have $N(A_1), ..., N(A_n)$ are independent.
- $N(\cdot)(\omega)$ is a measure on $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$, for all $\omega \in \Omega$.

This definition can be generalized from $(\mathbb{R}^d \setminus \{0\}, \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ to some measure space $(E, \mathcal{B}(E))$. Consider the following lemma about association of Poisson integrals.

Lemma 2.3.1. *Let N be a Poisson random measure with intensity measure ν. Let $g : \mathbb{R}^d \to \mathbb{R}^d_+$ and $g = (g_1, ..., g_d)$, with $\int_{\mathbb{R}^d} g_i(x)N(dx) < \infty$ for all $i = 1, ..., d$. Then*

$$\int_{\mathbb{R}^d} g(x)N(dx)$$

is an associated random vector in $\mathbb{R}^d$.

Proof. Let's begin with simple functions g. So $g(x) = \sum_{i=1}^n c_i \mathbb{1}_{A_i}(x)$, where $c_i \in \mathbb{R}^d_+$, $\{A_i\}$ disjoint. Then

$$\begin{aligned}\int_{\mathbb{R}^d} g(x)N(dx) &= \sum_{i=1}^n c_i N(A_i) \\ &= G(N(A_1), ..., N(A_n))\end{aligned}$$

where $(N(A_1), ..., N(A_n))$ is an associated random vector in $\mathbb{R}^n$, since $\{N(A_i)\}_{i=1}^n$ are independent, by Proposition 2.1.1 (g). Since $G : \mathbb{R}^n \to \mathbb{R}^d$ defined by $G(x_1, ..., x_n) = \sum_{i=1}^n c_i x_i$ is a non-decreasing function, then $G(N(A_1), ..., N(A_n))$ is associated by Proposition 2.1.1 (a), thus making $\int_{\mathbb{R}^d} g(x)N(dx)$ associated.

Now let g be any function $\mathbb{R}^d$ to $\mathbb{R}^d_+$ such that $\int_{\mathbb{R}^d} g_i(x)N(dx) < \infty$, a.s. for all $i = 1, ..., d$. Then g can be monotonically approximated by the sequence $\{g^{(n)}\}$ of positive simple functions, i.e. $g^{(n)} \nearrow g$ a.e. Then by monotone convergence theorem,

$$\int_{\mathbb{R}^d} g^{(n)}(x)N(dx) \to \int_{\mathbb{R}^d} g(x)N(dx) \qquad \text{a.s. } (\omega)$$

Thus, convergence also occurs weakly, and by Proposition 2.1.1 (d), we have $\int_{\mathbb{R}^d} g(x)N(dx)$ is an associated random vector.

The same result holds for "negative" functions, i.e. when $g : \mathbb{R}^d \to \mathbb{R}^d_-$. We could again use simple functions first, and we would get that our Poisson integral is a non-increasing function of an associated random vector, which would also make it associated. □

Theorem 2.7 (Resnick, 1988, [39]). *Let $X \sim ID(b, 0, \nu)$. If*

$$\nu((\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0, \tag{2.21}$$

i.e. ν concentrated on the positive and negative orthants $\mathbb{R}^d_+$ and $\mathbb{R}^d_-$, then X is associated.

Proof. We give a proof of Resnick's theorem using the Lévy-Itô decomposition. WLOG, assume $b = 0$. By Lévy-Ito decomposition, we have

$$X = \int_{|x|<1} x\tilde{N}(dx) + \int_{|x|\geq 1} xN(dx)$$

where N is a PRM of intensity measure ν. Let's first consider $\int_{|x|\geq 1} xN(dx)$. By (2.21), we can consider just integrating over the set

$$[\{|x| \geq 1\} \cap \mathbb{R}^d_+] \cup [\{|x| \geq 1\} \cap \mathbb{R}^d_-] =: A_1 \cup A_2$$

where A_i's are clearly disjoint. Observe that

$$\int_{A_1} xN(dx) = \int_{\mathbb{R}^d} h(x)N(dx),$$

where $h(x) = x\mathbb{1}_{A_1}(x) \in \mathbb{R}^d_+$. The components of the above integral are also finite a.s. from Lévy-Ito decomposition, and hence, by Lemma 2.3.1, $\int_{A_1} xN(dx)$ is associated. Also,

$$\int_{A_2} xN(dx) = \int_{\mathbb{R}^d} g(x)N(dx)$$

where $g(x) = x\mathbb{1}_{A_2}(x) \in \mathbb{R}^d_-$, and we have finiteness of components of integral. Thus Lemma 2.3.1 gives us $\int_{A_2} xN(dx)$ is associated. By disjointness of A_i's, we have independence between the Poisson integrals over A_1 and A_2. Therefore, the sum

$$\int_{|x|\geq 1} xN(dx) = \int_{A_1} xN(dx) + \int_{A_2} xN(dx)$$

is associated.

Now let's analyze $\int_{|x|<1} x\tilde{N}(dx)$. First consider $A_n = \{\frac{1}{n} < |x| < 1\}$ for all $n \in \mathbb{N}$. Then each set is bounded below, and by Lemma 2.3.4 in Applebaum [2], we have $N(A_n) < \infty$ a.s. and $\nu(A_n) < \infty$ for all n. Observe that, by (2.21), we can just integrate over

$$[A_n \cap \mathbb{R}^d_+] \cup [A_n \cap \mathbb{R}^d_-] = A_{n,1} \cup A_{n,2}.$$

By the same argument used above, $\int_{A_{n,j}} xN(dx)$ is associated for $j = 1, 2$. Also note that

$$\int_{A_{n,j}} x\nu(dx) \leq \int_{A_{n,j}} 1\nu(dx) \leq \nu(A_n) < \infty.$$

Then define $c_{n,j} = \int_{A_{n,j}} x\nu(dx)$. Then

$$\begin{aligned}
\int_{A_{n,j}} x\tilde{N}(dx) &= \int_{A_{n,j}} xN(dx) - \int_{A_{n,j}} x\nu(dx) \\
&= \int_{A_{n,j}} xN(dx) - c_{n,j} \\
&= G_j\left(\int_{A_{n,j}} xN(dx)\right)
\end{aligned}$$

where $G_j(x) = x - c_{n,j}$ in $\mathbb{R}^d$ for $j = 1, 2$. Since G_j is a nondecreasing function, we have $\int_{A_{n,j}} x\tilde{N}(dx)$ is associated. Thus

$$\int_{A_n} x\tilde{N}(dx) = \int_{A_{n,1}} x\tilde{N}(dx) + \int_{A_{n,2}} x\tilde{N}(dx)$$

is associated for all n, since the Poisson integrals on the RHS are independent. Finally,

$$\int_{A_n} x\tilde{N}(dx) \xrightarrow{L^2} \int_{|x|<1} x\tilde{N}(dx)$$

which implies convergence weakly. Therefore, $\int_{|x|<1} x\tilde{N}(dx)$ is associated by Proposition 2.1.1 (d), and

$$X = \int_{|x|<1} x\tilde{N}(dx) + \int_{|x|\geq 1} xN(dx)$$

is associated by, again, independence of the Poisson integrals. □

Condition (2.21) is only a sufficient condition. For a counter-example as to why it is not necessary, see Samorodnitsky [43]. There are interesting scenarios that arise in certain stochastic processes that yield (2.21) as sufficient and necessary. We will explore those in the next chapter.

We mentioned earlier that, unlike Gaussian distributions, for Poissonian distributions, the notion of association is strictly stronger than positive correlation. We present the following example.

Example 2.3.4. Let X be an α-stable distribution in $\mathbb{R}^d$, i.e. X arises as the weak limit of

$$\frac{Y_1 + ... + Y_n - b_n}{\sigma n^{1/\alpha}} \tag{2.22}$$

where Y_j are iid random vectors, $b_n \in \mathbb{R}^d$, $\sigma > 0$, and $\alpha \in (0, 2]$. When $\alpha = 2$, (2.22) will converge to a Gaussian distribution X by the classic central limit theorem. So we assume $\alpha < 2$. X is infinitely divisible with characteristics $(b, 0, \nu)$, where

$$\nu(B) = \int_{S^d} \int_0^\infty \mathbb{1}_B(r\theta) \frac{dr}{r^{1+\alpha}} \rho(d\theta) \tag{2.23}$$

where ρ is a finite Borel measure on the unit sphere S^d that describes the skewness and scale of the distribution and is called the **spectral measure**. These distributions have "heavy tails," where their decay in the tail is polynomially fast to 0, i.e.

$$f_X(x) \sim \frac{C}{|x|^{d+\alpha}}, \quad |x| \to \pm\infty \tag{2.24}$$

See Figure 2.2. The heavy-tail behavior causes $Var(X_i) = \infty$ or DNE, which in turn yields $\text{Cov}(X_i, X_j) = \infty$ or DNE. Hence, X cannot be PC. However, X can be associated. If we consider X such that ν is concentrated on $\mathbb{R}^d_+$ and $\mathbb{R}^d_-$, then X will be associated by Theorem 2.7. This will correspond to the spectral measure ρ being concentrated on $S^d \cap \mathbb{R}^d_+$ and $S^d \cap \mathbb{R}^d_-$.

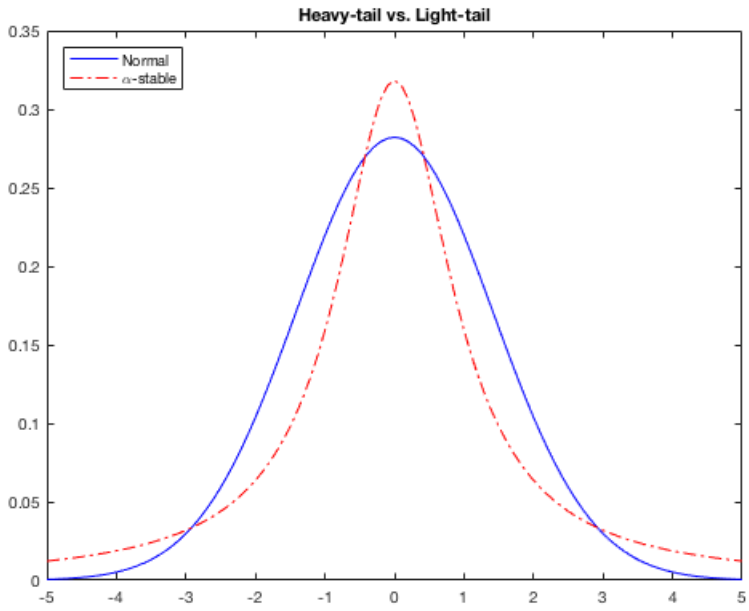

Figure 2.2: Gaussian vs. α-stable

2.3.3 Positive dependence and infinite divisibility of squared Gaussians

Very recently, the characterization of association for squared Gaussian random vectors has been given. Evans (1991) had posed a question about the characterization of association for such multivariate distributions. The recent result by Natalie Eisenbaum has provided a nice characterization, and also made a connection to infinite divisibility.

Theorem 2.8 (Eisenbaum, [17]). *Let $X = (X_1, ..., X_d)$ be a Gaussian random vector with mean 0. Then $X^2 := (X_1^2, ..., X_d^2)$ is associated if and only if it is infinitely divisible.*

Chapter 3

Positive dependence in Lévy-type Markov processes

In this chapter, we discuss the positive dependence of a certain class of time-homogeneous Markov processes. These processes will have behavior that is globally or locally "infinitely divisible-like." The Markov processes that exhibit this behavior are often called Lévy-type Markov processes. We will discuss Lévy-type Markov processes, how to analyze them, and how to describe their dependence structures. Our standard references here will be [8], [2]. For list of interesting applications of Lévy-type Markov processes in physics and finance, see [6].

3.1 Feller processes

Consider an adapted stochastic process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P}^x)_{x\in\mathbb{R}^d}$. $(\mathcal{G}_t)_{t\geq 0}$ is the filtration, and the index "x" indicates the starting point of the process: $\mathbb{P}^x(X_0 = x) = 1$, for all $x \in \mathbb{R}^d$. X is a **Markov process** if it satisfies the **Markov property**,

$$\mathbb{P}^x(X_t \in A|\mathcal{G}_s) = \mathbb{P}^x(X_t \in A|X_s), \quad \mathbb{P}^x\text{-a.s.} \tag{3.1}$$

for all $0 \leq s < t$, $A \in \mathcal{B}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$. Markov processes can be described by a family of linear operators $(T_{s,t})_{0 \leq s \leq t < \infty}$ on the Banach space $(B_b(\mathbb{R}^d), ||\cdot||_\infty)$ defined by

$$T_{s,t}f(x) = \mathbb{E}(f(X_t)|X_s = x). \tag{3.2}$$

We say Markov process X is **normal** if $T_{s,t} : B_b(\mathbb{R}^d) \to B_b(\mathbb{R}^d)$ for all $0 \leq s \leq t < \infty$.

Proposition 3.1.1. *A normal Markov process X with Markov evolution $(T_{s,t})_{0 \leq s \leq t < \infty}$ satisfies the following properties:*

1. $T_{s,s} = I$ *for all* $s \geq 0$.
2. $T_{r,s}T_{s,t} = T_{r,t}$, *for all* $0 \leq r \leq s \leq t < \infty$ *(Chapman-Kolmogorov).*
3. $f \geq 0$ *implies* $T_{s,t}f \geq 0$ *for all* $0 \leq s \leq t < \infty$ *(positivity-preserving).*
4. $||T_{s,t}|| \leq 1$ *for all* $0 \leq s \leq t < \infty$ *(contraction).*
5. $T_{s,t}1 = 1$.

Proof. See Applebaum [2]. □

Our focus in this chapter will be on **time-homogeneous Markov processes**, i.e. $T_{s,t} = T_{t-s}$. Thus, the Markov evolution can be described as a one-parameter, positivity-preserving contraction semigroup of operators $(T_t)_{t \geq 0}$. The Chapman-Kolmogorov equation can now be expressed as $T_sT_t = T_{s+t}$, which is also called the semigroup property. Thus we call $(T_t)_{t \geq 0}$ the **transition semigroup**.

Consider the Banach space of functions $(C_0(\mathbb{R}^d), ||\cdot||_\infty)$, where $C_0(\mathbb{R}^d)$ are continuous, bounded functions that vanish at infinity. We define the **infinitesimal generator** or **generator** $\mathcal{A}$ of process X to be an operator on $C_0(\mathbb{R}^d)$ by

$$\mathcal{A}f := \lim_{t \searrow 0} \frac{T_tf - f}{t}. \tag{3.3}$$

where the limit is with respect to the norm $||\cdot||_\infty$, i.e. uniform limit. The **domain of generator** $\mathcal{A}$, denoted $\mathcal{D}(\mathcal{A})$ is given by

$$\mathcal{D}(\mathcal{A}) = \left\{ f \in C_0(\mathbb{R}^d) : \lim_{t \searrow 0} \frac{T_t f - f}{t} \text{ exists uniformly} \right\}. \tag{3.4}$$

We will often write "generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$" to signify a generator and its domain.

A Markov process is called a **Feller process** if its semigroup $(T_t)_{t\geq 0}$ satisfies the **Feller property**, i.e. satisfies

(i) For all $t \geq 0$, $T_t : C_0(\mathbb{R}^d) \to C_0(\mathbb{R}^d)$ $\quad$ ($C_0(\mathbb{R}^d)$-invariant)

(ii) $\lim_{t\to 0} ||T_t f - f||_\infty = 0$, for all $f \in C_0(\mathbb{R}^d)$ $\quad$ (strong-continuity).

Feller semigroups satisfy a definition in functional analysis called C_0**-semigroups** or **strongly continuous one-parameter semigroups**, defined to be strongly-continuous, one-parameter semigroup of contractions on a Banach space $\mathbb{B}$. Such semigroups and their generators have nice properties with respect to the Banach spaces they live on. We will state these properties in terms of our Feller semigroup and generator on the Banach space $(C_0(\mathbb{R}^d), ||\cdot||_\infty)$.

Proposition 3.1.2. *Let X be a Feller process with semigroup $(T_t)_{t\geq 0}$ and generator $\mathcal{A}$.*

1. *$\mathcal{D}(\mathcal{A})$ is dense in $C_0(\mathbb{R}^d)$.*
2. *$T_t : \mathcal{D}(\mathcal{A}) \to \mathcal{D}(\mathcal{A})$, for all $t \geq 0$.*
3. *$\dfrac{d}{dt} T_t f = T_t \mathcal{A} f = \mathcal{A} T_t f$, for all $t \geq 0$, $f \in \mathcal{D}(\mathcal{A})$.*
4. *$\mathcal{A}$ is a closed operator.*

The derivative in Proposition 3.1.2(3) is defined based on the uniform limit with respect to $||\cdot||_\infty$. For a proof of the above proposition, see [2]. For a Feller generator $\mathcal{A}$, when $\mathcal{D}(\mathcal{A}) \supseteq C_c^\infty(\mathbb{R}^d)$, the space of infinitely differentiable functions with compact support, we call the process **rich Feller**. This is not a restrictive assumption, as almost all Feller processes that are studied satisfy this condition. The Feller generator $\mathcal{A}$ can take on nicer forms than the definition given in (3.3). We take an aside to discuss this.

Definition 3.1. An operator $p(x, D)$ on $S(\mathbb{R}^d)$, the Schwartz space, is a **pseudo-differential operator** if it has the form

$$p(x, D)f(x) = -(2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{ix\cdot\xi} p(x, \xi)\hat{f}(\xi)d\xi \tag{3.5}$$

where $\hat{f} \in S(\mathbb{R}^d)$ is the Fourier transform of f, and $p : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ is locally bounded in both arguments, $p(\cdot, \xi)$ measurable for all ξ, and $-p(x, \cdot)$ is a continuous negative definite function for all x.

According to Theorem 2.4, the function $p(x, \xi)$ will have a Lévy-Khintchine form for each x, i.e. there exist a triplet $(b(x), \Sigma(x), \nu(x, dy))$ such that

$$p(x, \xi) = ib(x) \cdot \xi - \frac{1}{2}\xi \cdot \Sigma(x)\xi + \int_{y\neq 0} (e^{i\xi\cdot y} - 1 - i\xi \cdot y\chi(y))\nu(x, dy). \tag{3.6}$$

where $b(x) \in \mathbb{R}^d$, $\Sigma(x) \in \mathbb{R}^{d\times d}$ symmetric positive semidefinite, and $\nu(x, \cdot)$ a Lévy measure, for each $x \in \mathbb{R}^d$. Analogous to Lévy-Khintchine, we will call $p(x, \xi)$ the **symbol** of the operator. We claim that the generator of a rich Feller process can take on this form.

Given a generator $\mathcal{A}$ of a C_0-semigroup $(T_t)_{t\geq 0}$ on a Banach space $\mathbb{B}$, the **resolvent** $R_\lambda(\mathcal{A})$ is defined to be

$$R_\lambda(\mathcal{A}) := (\lambda - \mathcal{A})^{-1}. \tag{3.7}$$

The **resolvent set**, $\rho(\mathcal{A})$, is the set of all λ such that the RHS of (3.7) exists. Note that $(0, \infty) \subseteq \rho(\mathcal{A})$, and for all $\lambda > 0$, $R_\lambda(\mathcal{A}) = \int_0^\infty e^{-\lambda t} T_t \, dt$, see [2]. The resolvent can easily be applied to the Feller generator and Feller semigroup on Banach space $(C_0(\mathbb{R}^d), ||\cdot||_\infty)$.

A linear operator A with domain $D(A)$ on $(C_0(\mathbb{R}^d), ||\cdot||_\infty)$ satisfies the **positive maximum principle (PMP)** if, for $f \in D(A)$, $\exists\, y_0 \in \mathbb{R}^d$ such that $f(y_0) = \sup_{y\in\mathbb{R}^d} f(y) \geq 0$, then $Af(y_0) \leq 0$. The resolvent set and PMP arise in an important theorem that characterizes generators of C_0-semigroups:

Theorem 3.1 (Hille-Yosida-Ray)**.** *Let A be a closed linear operator with domain $D(A)$ dense in $C_0(\mathbb{R}^d)$. Then A is the generator of a positivity-preserving strongly continuous contraction semigroup on $C_0(\mathbb{R}^d)$ if and only if*

- $(0, \infty) \subseteq \rho(A)$
- *A satisfies PMP.*

Proof. See Ethier and Kurtz [18]. □

Now comes the remarkable theorem of Courrège which connects PMP to pseudo-differential operators.

Theorem 3.2 (Courrège, [14])**.** *Suppose A is a linear operator on $C_0(\mathbb{R}^d)$ such that $A : C_c^\infty(\mathbb{R}^d) \to C(\mathbb{R}^d)$ and satisfies the positive maximum principle. Then $-A|_{C_c^\infty(\mathbb{R}^d)}$ is a pseudo-differential operator.*

Now let's return to the language of Feller processes.

Theorem 3.3. *For a Feller process X with generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ and semigroup $(T_t)_{t\geq 0}$, generator $\mathcal{A}$ satisfies the positive maximum principle.*

Proof. See Applebaum [2]. □

Corollary 3.1.1. *Let X be a Feller process with generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ such that $\mathcal{D}(\mathcal{A}) \supseteq C_c^\infty(\mathbb{R}^d)$, i.e. rich Feller. Then $-\mathcal{A}|_{C_c^\infty(\mathbb{R}^d)}$ is a pseudo-differential operator and takes on the form:*

$$\mathcal{A}|_{C_c^\infty(\mathbb{R}^d)} f(x) = -p(x, D)f(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{ix\cdot\xi} p(x, \xi)\hat{f}(\xi)d\xi. \tag{3.8}$$

*We call $p(x, \xi)$ the **symbol of the process**. The function $-p(x, \xi)$ is a cndf.*

Proof. Since generator $\mathcal{A}$ maps $\mathcal{D}(\mathcal{A})$ to $C_0(\mathbb{R}^d)$, the Banach space, then $\mathcal{A} : C_c^\infty(\mathbb{R}^d) \to C(\mathbb{R}^d)$. By Theorem 3.3, $\mathcal{A}$ satisfies positive maximum principle, and therefore, by Theorem 3.2, we have our desired result. □

The symbol can also be viewed in the sense of Jacob and Schilling. They give a more general definition of the symbol of the process $X = (X_t)_{t\geq 0}$:

$$\eta(x, \xi) = \lim_{t \searrow 0} \frac{\mathbb{E}^x(e^{i(X_t - x)\cdot\xi}) - 1}{t} \tag{3.9}$$

For rich Feller processes X, the RHS of (3.9) is exactly equal to the symbol $p(x,\xi)$ in the (3.8). Since it is a cndf, it will take on the Lévy-Khintchine form in (3.6). See [8] for more on this symbol. What's special about (3.9) is that it has been extended to more general Markov processes beyond Feller processes. Schnurr has shown that the symbol $\eta(x,\xi)$ in (3.9) is also defined for Lévy-Itô processes and homogeneous diffusions with jumps (in the sense of Jacod and Shiryaev [26]). He showed that $\eta(x,\xi)$ will also take on a Lévy-Khintchine form with characteristic triplet $(b(x),\Sigma(x),\nu(x,dy))$ corresponding to the differential semimartingale characteristics. Hence the symbol $\eta(x,\xi)$ can be used to study the behavior of the process, such as path properties, invariant measures, Haussdorf dimension, and more, for more general Markov processes, see [47], [48], [8]. We will maintain our focus on Feller processes.

3.1.1 Lévy processes

Lévy processes are examples of Feller processes.

Definition 3.2. A process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P}^x)_{x\in\mathbb{R}^d}$ is a **Lévy process** if is satisfies

- $X_0 = 0$ a.s.
- X_{t-s} independent of X_{s-r} for all $0 \leq r \leq s \leq t < \infty$ (independent increments)
- $X_t - X_s \stackrel{d}{=} X_{t-s}$ for all $0 \leq s \leq t$ (stationary increments)
- $\lim_{t\searrow 0} \mathbb{P}^x(|X_t - x| > a) = 0$ for all $a > 0$, for all $x \in \mathbb{R}^d$ (stochastic continuity)

The independent and stationary increments of a Lévy process yield a nice property: each X_t in a Lévy process is an infinitely divisible random vector. Additionally, the characteristic function of X_t looks like

$$\phi_{X_t}(\xi) = e^{t\psi(\xi)} \tag{3.10}$$

where ψ is the Lévy symbol of X_1, having a characteristic triplet (b,Σ,ν) and Lévy-Khintchine formula as given in (2.17). Thus, the behavior of the entire Lévy process can be characterized by the Lévy symbol ψ. Additionally, there is a 1-1 correspondence between cndf and Lévy processes, as signified in Theorem 2.4.

For a Lévy process X, the corresponding semigroup $(T_t)_{t\geq 0}$ has representation

$$T_t f(x) = \int_{\mathbb{R}^d} f(x+y)\mu_t(dy) = \mathbb{E}^0(f(X_t)+x) \tag{3.11}$$

where $f \in B_b(\mathbb{R}^d)$, $X_t \sim \mu_t$, for all $t \geq 0$. This representation yields the classification of Lévy processes as the subclass of Markov processes which are **translation invariant** [2]. The generator $\mathcal{A}$ of X is of course also a pseudo-differential operator on the Schwartz space $S(\mathbb{R}^d)$ which contains $C_c^\infty(\mathbb{R}^d)$.

Theorem 3.4. *Let X be a Lévy process with generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ and Lévy symbol*

$$\psi(\xi) = i\xi\cdot b - \frac{1}{2}\xi\cdot\Sigma\xi + \int_{y\neq 0}(e^{i\xi\cdot y} - 1 - i\xi\cdot y\chi(y))\nu(dy).$$

Then $-\mathcal{A}|_{S(\mathbb{R}^d)}$ is a pseudo-differential operator and

$$\mathcal{A}|_{S(\mathbb{R}^d)} f(x) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} e^{ix\cdot\xi}\psi(\xi)\hat{f}(\xi)d\xi. \tag{3.12}$$

Proof. See Applebaum [2]. □

So the symbol of the process in (3.12) coincides with the Lévy symbol! Thus, when we say "symbol," we will be simultaneously referring to both of these objects. Additionally, we say (b, Σ, ν) is the characteristic triplet of the Lévy process. To interpret the meaning of the Lévy symbol and the symbol of the process for general Feller processes, we will look at the characteristic triplet and examples.

Example 3.1.1 (Brownian motion)**.** Let X be a Lévy process with symbol ψ defined by the characteristic triplet $(b, \Sigma, 0)$. Then

$$\phi_{X_t}(\xi) = \exp\left(t\left(i\xi\cdot b - \frac{1}{2}\xi\cdot\Sigma\xi\right)\right),$$

yielding $X_t \sim N(tb, t\Sigma)$. By independent and stationary increments and stochastic continuity of a Lévy process, we have that X is a Brownian motion with drift b and covariance Σ.

Example 3.1.2 (Poisson process)**.** Let X be a Lévy process on $\mathbb{R}$ with symbol ψ defined by the characteristic triplet $(0, 0, \lambda\delta_1)$. Then

$$\begin{aligned}\phi_{X_t}(\xi) &= \exp\left(t\left(\int_{y\neq 0}(e^{i\xi\cdot y} - 1 - i\xi\cdot y\chi(y))\lambda\delta_1(dy)\right)\right)\\ &= \exp(t\lambda(e^{i\xi} - 1))\end{aligned}$$

So $X_t \sim Poiss(\lambda t)$ for all $t \geq 0$, and by independent and stationary increments and stochastic continuity, X is a Poisson process. X has jump rate λ and with jump size 1, as given by δ_1.

Example 3.1.3 (Compound Poisson process)**.** Let X be a Lévy process on $\mathbb{R}$ with symbol ψ defined by the characteristic triplet $(b, 0, \lambda\mu)$, where $\lambda > 0$ and μ is a probability measure, and $b = \int y\chi(y)\lambda\mu(dy)$. Then

$$\begin{aligned}\phi_{X_t}(\xi) &= \exp\left(t\left(i\xi\cdot\int_{y\neq 0} y\chi(y)\lambda\mu(dy) + \int_{y\neq 0}(e^{i\xi\cdot y} - 1 - i\xi\cdot y\chi(y))\lambda\mu(dy)\right)\right)\\ &= \exp\left(t\left(\int_{y\neq 0}(e^{i\xi\cdot y} - 1)\lambda\mu(dy)\right)\right)\end{aligned}$$

Then $X_t = \sum_{i=1}^{N_t} Z_i$, where $Z_i \overset{iid}{\sim} \mu$, and $N = (N_t)_{t\geq 0}$ is Poisson process of rate λ. Thus, X is a compound Poisson process with jump rate λ and with jump size determined by the common law μ.

Thus, Lévy-Khintchine gives us a nice interpretation of the symbol and characteristic triplet of a Lévy process. b represents the non-random linear drift of the process, Σ represents the covariance of the Brownian motion component, and Lévy measure ν represents the intensity and size of the jumps of the process. In other words, b, Σ represent the continuous (Brownian motion with drift) behavior, and ν represents the jump (Poissonian) behavior. We call the jump behavior "Poissonian" because the jump part can be represented by a compound Poisson process or the L^2-limit of compound Poisson processes (analogous to the idea of Poissonian distributions which we mentioned after Theorem 2.16). Our interpretation can be explicitly supported by the Lévy-Itô decomposition of a Lévy process.

Theorem 3.5 (Lévy-Itô)**.** *Let X be a Lévy process with symbol ψ and characteristic triplet (b, Σ, ν). Then for all $t \geq 0$,*

$$X_t = bt + B_t + \int_{|y|<1} y\tilde{N}(t, dy) + \int_{|y|\geq 1} yN(t, dy), \tag{3.13}$$

where $B = (B_t)_{t\geq 0}$ is a mean-zero Brownian motion with covariance Σ, and $N : \mathbb{R}_+ \times \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ is a Poisson random measure with intensity $\mathbb{E}N(t, A) = t\nu(A)$, $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$, i.e. $N(t, A) \sim Poiss(t\nu(A))$.

Thus, a Lévy process is decomposed into the sum of two independent processes, a continuous Brownian motion with drift (1st two terms of (3.13)), and a jump process (Poisson integrals), where the first integral represents "small jumps" of jump-size less than 1, and the second integral represents "large jumps" of jump-size greater than or equal to 1.

Let's now return to general Feller processes. Observe that, in the symbol $p(x, \xi)$ of a Feller process, the characteristic triplet that defines the symbol depends on variable x: $(b(x), \Sigma(x), \nu(x, dy))$. This x value corresponds to the starting point of the process, i.e. $\mathbb{P}^x(X_0 = x) = 1$. Thus, these parameters $b(x), \Sigma(x), \nu(x, dy)$ depend on the starting location of the process, and thus signifies that their corresponding Feller process is state-space dependent. This corresponds nicely to the interpretation of Lévy processes which have constant characteristic triplets (b, Σ, ν). No dependence on x indicates no dependence on the state space, which corresponds to the notion of Lévy processes being a subclass of Markov processes which are homogeneous in space.

Recall that in Lévy processes, b, Σ characterize the continuous component (Brownian motion with drift) of the process and ν characterizes the jump component (Poissonian) of the process. Now for a general Feller process with characteristics $(b(x), \Sigma(x), \nu(x, dy))$, functions $b(x), \Sigma(x)$ still describe the continuous behavior of the process, and $\nu(x, dy)$ still describes the jump behavior. However, unlike the Lévy case, these characteristics now depend on the state x. Moreover, the $b(x), \Sigma(x)$ cannot be interpreted *globally* as Brownian motion with drift, and $\nu(x, dy)$ cannot be interpreted *globally* as mean intensity of a Poissonian process. They can however have a *local* interpretation. In other words, let X be a Feller process with characteristics $(b(x), Q(x), \nu(x, dy)$ and fix a starting point $x_0 \in \mathbb{R}^d$, i.e. $\mathbb{P}^{x_0}(X_0 = x_0) = 1$.

Then in short time, the behavior of X can be approximated by a process $(Y_t + x_0)_{t\geq 0}$, where Y is a Lévy process with characteristics $(b(x_0), Q(x_0), \nu(x_0, dy))$. This interpretation comes from Jacob and Schilling's definition of the symbol in (3.9) and is discussed in more detail in [8]. This local-Lévy behavior of Feller processes also leads to the name of **Lévy-type Markov processes** that is sometimes given in literature in place of "Feller processes."

The interpretation of $b(x), \Sigma(x)$ and $\nu(x, dy)$ as representing the continuous and jump components, respectively, of the process can also be seen in the semimartingale characteristics and the semimartingale decomposition of a Feller process, which is more general than the Lévy-Itô decomposition. For more on this, see [47].

Remark 3.1.1. (i) Recall that Theorem 2.4 gave us a 1-1 correspondence between continuous negative definite functions and Lévy processes. In other words, if we start with a Lévy process, we can get obtain a symbol ψ characterizing the process, and if we start with a cndf $-\psi$, we obtain a corresponding Lévy process X with symbol ψ.

(ii) We do not have the same such correspondence between general Feller processes and symbols $p(x, \xi)$. The theorem of Courrege (Theorem 3.2) gave us that if we start with a Feller process, then we can get a symbol $p(x, \xi)$. But if we begin with a function $-p(x, \xi)$ that is cndf in ξ and has dependence on x, then we may not necessarily be able to construct a Feller process from that. We can construct a pseudo-differential operator $p(x, D)$, and such an operator satisfies positive maximum principle. However, in order to generate a Feller semigroup, we also need the first condition of Hille-Yosida-Ray theorem (Theorem 3.1) to be satisfied. For more on conditions for symbols to construct Feller processes, see Walter Hoh's thesis [24].

(iii) The characteristics $(b(x), \Sigma(x), \nu(x, dy))$ and symbol $p(x, \xi)$ characterize the behavior of the Feller process. They are often used to analyze various properties of the process. We will use these parameters to characterize the dependence properties of the process.

3.2 Positive dependence in general Markov processes

We discussed in Chapter 2 different forms of positive dependence in multivariate distributions. Now we discuss what positive dependence means for stochastic processes. We will have two different notions.

Definition 3.3. A stochastic process $X = (X_t)_{t\geq 0}$ in $\mathbb{R}^d$ is **associated in time** or **temporally associated** if, for all $0 \leq t_1 < ... < t_n$, the random vector $(X_{t_1}, ..., X_{t_n})$ is associated in $\mathbb{R}^{dn}$.

Definition 3.4. A stochastic process $X = (X_t)_{t\geq 0}$ in $\mathbb{R}^d$ is **associated in space** or **spatially associated** if, for all $t \geq 0$, the random vector X_t is associated in $\mathbb{R}^d$.

We can define WA, PSA, PSD, POD, PUOD, and PLOD in time and space for a stochastic process as well by replacing "associated" in the above definitions with "WA," "PSA," "PSD," "POD," "PUOD," and "PLOD." Clearly, Definition 3.3 implies Definition 3.4. The rough interpretation of spatial dependence in a stochastic process is that the components of the process $X_t^{(1)}, ..., X_t^{(d)}$ move, on average, in the "same direction." Temporal dependence is even stronger than this notion, in that occurrences at moments X_t effect future occurrences X_{t+s} in a positively dependent way.

There are certain cases when temporal and spatial dependence are equivalent. Lévy processes are such an example.

Theorem 3.6. *Let $X = (X_t)_{t\geq 0}$ be a stochastic process in $\mathbb{R}^d$ with independent and stationary increments. Then X is associated in time if and only if X is associated in space.*

Proof. ($\Rightarrow$) Trivial.

($\Leftarrow$) Assume X_t is associated in $\mathbb{R}^d$ for every $t \geq 0$. Choose $0 \leq t_1 < ... < t_n$. Then

$$\begin{aligned}(X_{t_1}, ..., X_{t_n}) &= (X_{t_1}, X_{t_1} + (X_{t_2} - X_{t_1}), ..., X_{t_1} + (X_{t_2} - X_{t_1}) + ... + (X_{t_n} - X_{t_{n-1}})) \\ &= (X_{t_1}, ..., X_{t_1}) + (\vec{0}, X_{t_2} - X_{t_1}, ..., X_{t_2} - X_{t_1}) + ... + (\vec{0}, ..., \vec{0}, X_{t_n} - X_{t_{n-1}}).\end{aligned}$$

Now observe that by stationary increments, $X_{t_{k+1}} - X_{t_k} \stackrel{d}{=} X_{t_{k+1}-t_k}$ and $X_{t_{k+1}-t_k}$ is associated, which makes $X_{t_{k+1}} - X_{t_k}$ associated (association is preserved under equality in distribution),

for all $k \in \{1, ..., n-1\}$. Now observe that if $\hat{X}$ is associated in $\mathbb{R}^d$, then each block $(\vec{0}, ...\vec{0}, \hat{X}, ..., \hat{X})$ is associated in $\mathbb{R}^{dn}$, where there are k number of $\vec{0}$ vectors and $(n-k)$ $\hat{X}$ vectors. Therefore, each block $(\vec{0}, ..., \vec{0}, X_{t_{k+1}} - X_{t_k}, ..., X_{t_{k+1}} - X_{t_k})$ is associated, for each $k \in \{1, ..., n-1\}$. By independent increments, each block is independent. Therefore, since the sum of independent random vectors, each of which is associated, is associated, then $(X_{t_1}, ..., X_{t_n})$ is associated. □

Corollary 3.2.1. *A Lévy process X that is associated in space is also associated in time.*

Remark 3.2.1. (i) The above theorem holds with "associated" replaced by "PSD," "POD," "PUOD," and "PLOD"

(ii) The idea behind the result in Theorem 3.6 comes from Bauerle [4], whom claimed that Theorem 3.6 was true for just processes with independent increments, dropping the assumption of stationary increments. But this result fails without the assumption of stationary increments. To see this, one can consider a general process with independent increments in the sense of Jacod and Shiryaev, or an additive process.

In time-homogeneous Markov processes, additional conditions can yield spatial association implying temporal association. We investigate this for Markov processes that are stochastically monotone:

Definition 3.5. A Markov process $X = (X_t)_{t \geq 0}$ with semigroup $(T_t)_{t \geq 0}$ is **stochastically monotone** if $f \in \mathcal{F}_i$ implies $T_t f \in \mathcal{F}_i$, for all $t \geq 0$.

We can state stochastic monotonicity with respect to different functions classes, like $\mathcal{F}_{ism}$. For this chapter, we will only focus on the one given with respect to $\mathcal{F}_i$ as given in Definition 3.5. The idea of stochastic monotonicity is much related to stochastic orders $\mathcal{F}$ as given in Chapter 2. For more on this subject, see [34].

Another interpretation of spatial association is that the "process preserves positive correlations," as was given by Liggett [32] and Mu Fa Chen [11]. We describe this notion here and prove that it is equivalent to our definition of spatial association.

Let $\mathcal{M}_a$ be the space of probability measures μ on $\mathbb{R}^d$ such that

$$\int_{\mathbb{R}^d} fg \, d\mu \geq \int_{\mathbb{R}^d} f \, d\mu \int_{\mathbb{R}^d} g \, d\mu, \quad \forall f, g \in \mathcal{F}_i, \tag{3.14}$$

where the integrals are finite. We call $\mathcal{M}_a$ the **space of associated measures**. We can consider a space $\mathcal{M}_a$ of measures on any particular state space E with partial order, not just $\mathbb{R}^d$. Given a Markov process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{G}, (\mathcal{G})_{t\geq 0}, \mathbb{P}^x)_{x\in\mathbb{R}^d}$ with semigroup $(T_t)_{t\geq 0}$, if $X_0 \sim \mu$, we define a probability measure $\mathbb{P}^\mu$ by the following:

$$\mathbb{E}^\mu f(X_t) = \int_{\mathbb{R}^d} T_t f(y)\mu(dy) \tag{3.15}$$

Additionally, for any probability measure μ, where $X_0 \sim \mu$, we can define the probability measure μT_t by the prescription

$$\int_{\mathbb{R}^d} f(y)\mu T_t(dy) := \int_{\mathbb{R}^d} T_t f(y)\mu(dy) = \mathbb{E}^\mu f(X_t), \quad f \in B_b(\mathbb{R}^d). \tag{3.16}$$

So we say Markov process X **preserves positive correlations** if $\mu \in \mathcal{M}_a$ implies $\mu T_t \in \mathcal{M}_a$ for all $t \geq 0$.

Lemma 3.2.1. *Let $X = (X_t)_{t\geq 0}$ is a Markov process on $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P}^x)_{x\in\mathbb{R}^d}$, and assume further that X is stochastically monotone.*
Statement 1: X_t associated for all $t \geq 0$ with respect to $\mathbb{P}^x$, for all $x \in \mathbb{R}^d$, i.e.

$$T_t fg(x) \geq T_t f(x) T_t g(x) \quad \forall x \in \mathbb{R}^d.$$

Statement 2: If $X_0 \sim \mu \in \mathcal{M}_a$, then $\mu T_t \in \mathcal{M}_a$, for all $t \geq 0$.
Then we have Statement 1 $\iff$ Statement 2.

Proof. ($\Leftarrow$). Choose $x \in \mathbb{R}^d$, and let $X_0 \sim \mu = \delta_x$. Observe that $\delta_x \in \mathcal{M}_a$, since

$$\int fg(y)\delta_x(dy) = f(x)g(x) = \int f(y)\delta_x(dy) \int g(y)\delta_x(dy).$$

Then $\mu T_t = \delta_x T_t \in \mathcal{M}_a$ for all $t \geq 0$, which implies

$$\begin{aligned} T_t fg(x) = \int T_t fg(y)\delta_x(dy) = \int fg(y)\delta_x T_t(dy) &\geq \int f(y)\delta_x T_t(dy) \int g(y)\delta_x T_t(dy) \\ &= \int T_t f(y)\delta_x(dy) \int T_t g(y)\delta_x(dy) \end{aligned}$$

$$= T_t f(x) T_t g(x)$$

($\Rightarrow$). Choose $\mu \in \mathcal{M}_a$.

$$\begin{aligned}\int f g(y) \mu T_t(dy) = \int T_t f g(y) \mu(dy) &\geq \int T_t f(y) T_t g(y) \mu(dy), \text{ by (1)} \\ &\geq \int T_t f(y) \mu(dy) \int T_t g(y) \mu(dy), \text{ by stoch. monotonicity} \\ &= \int f(y) \mu T_t(dy) \int g(y) \mu T_t(dy).\end{aligned}$$

So $\mu T_t \in \mathcal{M}_a$. □

Liggett gives a condition for stochastically monotone, spatially associated Markov processes to also be temporally associated [32]. His result is for Markov processes on compact state spaces with bounded generators, but it can be easily extended to more general Markov processes on $\mathbb{R}^d$. We present the statement and proof here.

Theorem 3.7 (Liggett [32]). *Let $X = (X_t)_{t \geq 0}$ be a time-homogeneous, stochastically monotone Markov process on $\mathbb{R}^d$ that is associated in space, i.e. X_t is associated for all $t \geq 0$ with respect to $\mathbb{P}^x$, for all $x \in \mathbb{R}^d$. Assume additionally that $X_0 \sim \mu$, where $\mu \in \mathcal{M}_a$. Then X is associated in time.*

Proof. Want to show that $(X_{t_1}, ..., X_{t_n})$ is associated with respect to $\mathbb{P} = \mathbb{P}^\mu$ where μ is the probability measure of X_0, i.e. $X_0 \sim \mu \in \mathcal{M}_a$. Our goal is to show

$$\mathbb{E}^\mu f(X_{t_1}, ..., X_{t_n}) g(X_{t_1}, ..., X_{t_n}) \geq \mathbb{E}^\mu f(X_{t_1}, ..., X_{t_n}) \mathbb{E}^\mu g(X_{t_1}, ..., X_{t_n}) \tag{3.17}$$

for all $f, g : \mathbb{R}^{dn} \to \mathbb{R}$. We do a proof by induction:

Base Case: $n = 1$. X_{t_1} is associated wrt $\mathbb{P}^x$ for all x by assumption. By that fact, stochastic monotonicity and $\mu \in \mathcal{M}_a$,

$$\begin{aligned}\mathbb{E}^\mu f g(X_{t_1}) = \int T_{t_1} f g(y) \mu(dy) \geq \int T_{t_1} f(y) T_{t_1} g(y) \mu(dy) &\geq \int T_{t_1} f(y) \mu(dy) \int T_{t_1} g(y) \mu(dy) \\ &= \mathbb{E}^\mu f(X_{t_1}) \mathbb{E}^\mu g(X_{t_1})\end{aligned}$$

Induction Hypothesis: $n-1$. Assume we have inequality (3.17) for $n-1$. Define $s_k = t_{k+1} - t_1$, $k \in \{1, ..., n-1\}$. Then for a given $x \in \mathbb{R}^d$, we would have

$$\mathbb{E}^\mu f(x, X_{s_1}, ..., X_{s_{n-1}}) g(x, X_{s_1}, ..., X_{s_{n-1}}) \geq \mathbb{E}^\mu f(x, X_{s_1}, ..., X_{s_{n-1}}) \mathbb{E}^\mu g(x, X_{s_1}, ..., X_{s_{n-1}})$$

where again $\mu \in \mathcal{M}_a$. So this inequality would also hold for $\mu = \delta_x \in \mathcal{M}_a$, i.e.

$$\mathbb{E}^x f(x, X_{s_1}, ..., X_{s_{n-1}}) g(x, X_{s_1}, ..., X_{s_{n-1}}) \geq \mathbb{E}^x f(x, X_{s_1}, ..., X_{s_{n-1}}) \mathbb{E}^x g(x, X_{s_1}, ..., X_{s_{n-1}}).$$

Inductive Step: n and $\mathbb{E} = \mathbb{E}^\mu$.

$$\begin{aligned}
&\mathbb{E} f(X_{t_1}, ..., X_{t_n}) g(X_{t_1}, ..., X_{t_n}) \\
&= \int \mathbb{E}(f(X_{t_1}, ..., X_{t_n}) g(X_{t_1}, ..., X_{t_n}) | X_{t_1} = x) \mu_{X_{t_1}}(dx) \\
&= \int \mathbb{E}(f(x, X_{t_2}, ..., X_{t_n}) g(x, X_{t_2}, ..., X_{t_n}) | X_{t_1} = x) \mu_{X_{t_1}}(dx) \\
&= \int \mathbb{E}(f(x, X_{t_2 - t_1}, ..., X_{t_n - t_1}) g(x, X_{t_2 - t_1}, ..., X_{t_n - t_1}) | X_0 = x) \mu_{X_{t_1}}(dx), \text{ by time-homo. M.P} \\
&= \int \mathbb{E}(f(x, X_{s_1}, ..., X_{s_{n-1}}) g(x, X_{s_1}, ..., X_{s_{n-1}}) | X_0 = x) \mu_{X_{t_1}}(dx) \\
&= \int \mathbb{E}^x f(x, X_{s_1}, ..., X_{s_{n-1}}) g(x, X_{s_1}, ..., X_{s_{n-1}}) \mu_{X_{t_1}}(dx) \\
&\geq \int \mathbb{E}^x f(x, X_{s_1}, ..., X_{s_{n-1}}) \mathbb{E}^x g(x, X_{s_1}, ..., X_{s_{n-1}}) \mu_{X_{t_1}}(dx) \\
&\geq \int \mathbb{E}^x f(x, X_{s_1}, ..., X_{s_{n-1}}) \mu_{X_{t_1}}(dx) \int \mathbb{E}^x g(x, X_{s_1}, ..., X_{s_{n-1}}) \mu_{X_{t_1}}(dx) \\
&= \mathbb{E} f(X_{t_1}, ..., X_{t_n}) \mathbb{E} g(X_{t_1}, ..., X_{t_n})
\end{aligned}$$

where the first inequality comes from Induction Hypothesis, and the second inequality comes from Base Case. □

Remark 3.2.2. The idea in Theorem 3.7 was first stated by Harris in for Markov processes on a finite, partially ordered state space [22].

3.3 Positive dependence in Lévy processes

Association of Lévy processes has been well characterized. The ideas behind the characterization based on the Lévy triplet (b, Q, ν) stem from the conditions for association in infinitely divisible distributions that were presented in Chapter 2. Since a Lévy process X can be decomposed into the independent sum of Brownian motion and a Poissonian jump process, we can study the characterization of association for Lévy processes by looking the cases $(0, \Sigma, 0)$ and $(b, 0, \nu)$ independently.

Theorem 3.8 (Herbst, Pitt, [23]). *Let $X = (X_t)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ with covariance Σ, i.e. a Lévy process with triplet $(0, \Sigma, 0)$. Then X is spatially associated if and only if $\Sigma_{ij} \geq 0$ for all $i, j \in \{1, ..., d\}$. Additionally, this is equivalent to temporal association by Corollary 3.2.1.*

Theorem 3.9 (Samorodnitsky, [43]). *Let $X = (X_t)_{t\geq 0}$ be a Lévy process in $\mathbb{R}^d$ with triplet $(b, 0, \nu)$. Then X is spatially associated if and only if ν is concentrated on the positive and negative orthants, i.e.*

$$\nu((\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0. \tag{3.18}$$

Additionally, this is equivalent to temporal association by Corollary 3.2.1.

3.4 Positive dependence in Feller processes

We want to now study positive dependence for more general Feller processes, in particular, stochastically monotone Feller processes. Mu-Fa Chen characterized association for stochastically monotone Feller processes with characteristics $(b(x), \Sigma(x), 0)$.

Theorem 3.10 (Mu-Fa Chen, [11]). *Let $X = (X_t)_{t\geq 0}$ be a stochastically monotone Feller process with triplet $(b(x), \Sigma(x), 0)$. Then X is spatially associated if and only if $\Sigma_{ij}(x) \geq 0$ for all $i, j \in \{1, ..., d\}$, $x \in \mathbb{R}^d$.*

Chen also provides sufficient conditions for stochastic monotonicity based on the parameters $b(x), \Sigma(x)$: $b_i(x)$ is an increasing function in component x_k, for all $i \neq k$, and $\Sigma_{ij}(x)$ only depends on components x_i, x_j. See [11].

Our interest lies in jump-Feller processes: $(b(x), 0, \nu(x, dy))$. Ideally, we want to extend the result of Samorodnitsky in Theorem 3.9 to jump-Feller processes. This idea was previously worked out by Jie-Ming Wang [52] (was unknown to the author at the time), but under many continuity and integrability conditions on the characteristics. The assumptions include

- $b_i, \Sigma_{ij} \in C(\mathbb{R}^d)$, for all i, j.
- $\int h_i(z)(\nu(\cdot, dz) - \nu(\cdot, d(-z))) \in C(\mathbb{R}^d)$, where $h : \mathbb{R}^d \to \mathbb{R}^d$ is defined by $h_i(z) = \text{sgn}(z_i)(1 \wedge |z_i|)$.
- $\int_A |h(z)|^2 \nu(\cdot, dz) \in C(\mathbb{R}^d)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$.
- $\int g(z)\nu(\cdot, dz) \in C(\mathbb{R}^d)$ for all $g \in C_b(\mathbb{R}^d)$ that is 0 near the origin.

We relax these conditions, and extend the result to the other forms of positive dependence mentioned in Chapter 2: WA, PSA, PSD, POD, PUOD and PLOD. Our main tools in this extension are quite different than those in Wang's proof and include the symbol, integro-differential operators, and small-time asymptotics of Feller processes.

3.4.1 Bounded symbols

As stated in Definition 3.1, the symbol $p(x, \xi)$ is **locally bounded**. This means for every $K \subset \mathbb{R}^d$ compact, there exists a constant $C_K > 0$ such that

$$\sup_{x \in K} |p(x, \xi)| \leq C_K(1 + |\xi|^2). \tag{3.19}$$

Being locally bounded in x is equivalent to local boundedness of characteristic triplet in the following sense:

Theorem 3.11 (Schilling, [46]). *Given a symbol $p(x, \xi)$ defined by the characteristics $(b(x), \Sigma(x), \nu(x, dy))$, then $p(x, \xi)$ is locally bounded if and only if*

$$||b||_{\infty,K} + ||\Sigma||_{\infty,K} + \left|\left|\int_{y \neq 0} (1 \wedge |y|^2)\nu(\cdot, dy)\right|\right|_{\infty,K} < \infty, \tag{3.20}$$

where $||f||_{\infty,K} := \sup_{x \in K} |f(x)|$, for every compact set $K \subset \mathbb{R}^d$.

We say a symbol $p(x,\xi)$ is **bounded** if we can replace K above with $\mathbb{R}^d$, i.e. there exists $C > 0$ s.t. $\sup_{x\in\mathbb{R}^d} |p(x,\xi)| \leq C(1+|\xi|^2)$. This is equivalent to boundedness of the characteristics in the sense of Theorem 3.11 with $||\cdot||_{\infty,K}$ replaced by $||\cdot||_\infty$, the sup-norm. Bounded symbols give us more analytical power, as we will see when we look at the integro-differential operator. We will often assume a bounded symbol throughout this chapter, but in several occasions, we will highlight where we don't require boundedness.

3.4.2 Integro-differential operator, extended generator

We know from Section 3.1 that the generators of rich Feller processes become pseudo-differential operators on the space $C_c^\infty(\mathbb{R}^d)$. We obtain the integro-differential operator by substituting the Lévy-Khintchine representation of the symbol

$$p(x,\xi) = ib(x)\cdot\xi - \frac{1}{2}\xi\cdot\Sigma(x)\xi + \int_{y\neq 0}(e^{i\xi\cdot y} - 1 - i\xi\cdot y\chi(y))\nu(x,dy) \tag{3.21}$$

into the RHS of the pseudo-differential operator

$$\mathcal{A}|_{C_c^\infty(\mathbb{R}^d)}f(x) = -p(x,D)f(x) = (2\pi)^{-d/2}\int_{\mathbb{R}^d} e^{ix\cdot\xi}p(x,\xi)\hat{f}(\xi)d\xi. \tag{3.22}$$

After some elementary Fourier analysis, one can obtain the integro-differential operator, which we denote by $I(p)$:

$$I(p)f(x) = b(x)\cdot\nabla f(x) + \frac{1}{2}\nabla\cdot\Sigma(x)\nabla f(x) + \int_{y\neq 0}\left(f(x+y) - f(x) - y\cdot\nabla f(x)\chi(y)\right)\nu(x,dy) \tag{3.23}$$

where $\nabla\cdot\Sigma(x)\nabla f(x) = \sum_{j,k=1}^d \Sigma_{jk}(x)\partial_j\partial_k f(x)$. For details on the Fourier analysis, see [2]. The RHS of $I(p)$ in (3.23) exists for $f \in C_b^2(\mathbb{R}^d)$, the space of continuous, bounded, twice-differentiable functions.

$I(p)$ is also considered an extended generator. Extended generators can be defined in various ways. But we consider the definition given by Schilling and Schnurr [47]. For a Markov process X with generator $\mathcal{A}$, we define the **extended domain** $\widetilde{\mathcal{D}(\mathcal{A})}$ to be the set

of $f \in B_b(\mathbb{R}^d)$ such that there exists $g \in B_b(\mathbb{R}^d)$,

$$f(X_t) - f(X_0) - \int_0^t g(X_s)ds \tag{3.24}$$

is a well-defined and also a local martingale with respect to $\mathbb{P}^x$ for all $x \in \mathbb{R}^d$. For all $f \in \widetilde{\mathcal{D}(\mathcal{A})}$, if we choose one $g \in B_b(\mathbb{R}^d)$ satisfying (3.24) and say $\widetilde{\mathcal{A}}f := g$. Then we say $\widetilde{\mathcal{A}}$ (is a version of) the **extended generator**.

Operator $I(p)$ of a Feller process is such an extended generator of $\mathcal{A}$ onto the extended domain $C_b^2(\mathbb{R}^d)$, as given by the next two results.

Theorem 3.12 (Schilling, [46]). *Let X be a rich Feller process with generator $\mathcal{A}$, bounded symbol $p(x,\xi)$, and integro-differential operator $(I(p), C_b^2(\mathbb{R}^d))$. Then $I(p)$ is the unique extension of $\mathcal{A}$ onto $C_b^2(\mathbb{R}^d)$ such that*

$$||I(p)u||_\infty \leq C \sum_{|\alpha|\leq 2} ||\partial^\alpha u||_\infty.$$

Thus, $I(p)|_{\mathcal{D}(\mathcal{A})} = \mathcal{A}$.

Corollary 3.4.1 (Schilling, [46]). *Let X be a rich Feller process with generator $\mathcal{A}$, bounded symbol $p(x,\xi)$, and integro-differential operator $(I(p), C_b^2(\mathbb{R}^d))$. The process*

$$M_t^f := f(X_t) - f(X_0) - \int_0^t I(p)f(X_{s-})ds$$

is, for every $f \in C_b^2(\mathbb{R}^d)$, an L^2-local martingale.

The above corollary will provide a useful way to use analyze the dependence properties of the Feller process, which we will see when we present our main results. Also, $(I(p), C_b^2(\mathbb{R}^d))$ provides a larger class of functions than the pseudo-differential operator $p(x,D)$ (on $C_c^\infty(\mathbb{R}^d)$) and the generator $\mathcal{A}$ (on $\mathcal{D}(\mathcal{A}) \subseteq C_0(\mathbb{R}^d)$). We desire such a larger class of functions, since we will want to look at monotone functions on $\mathbb{R}^d$ to characterize association.

To better see this, we present an important theorem of Liggett.

Theorem 3.13 (Liggett, [32]). *Let $X = (X_t)_{t\geq 0}$ be a Feller process on state space E with generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ and semigroup $(T_t)_{t\geq 0}$. If X is stochastically monotone, i.e. $T_t f \in \mathcal{F}_i$*

for all $f \in \mathcal{F}_i$, *then*

$$\mathcal{A}fg \geq g\mathcal{A}f + f\mathcal{A}g, \quad \forall f, g \in \mathcal{F}_i \cap \mathcal{D}(\mathcal{A}) \tag{3.25}$$

if and only if

$$\mu \in \mathcal{M}_a \implies \mu T_t \in \mathcal{M}_a, \quad \forall t \geq 0. \tag{3.26}$$

(Recall that (3.26) is equivalent to our notion of spatial association.) Liggett proved this for E compact and $\mathcal{A}$ bounded. This was extended by Szekli and Ruschendorf to more general Polish spaces E and $\mathcal{A}$ unbounded [49, 42]. For the Feller processes we consider, particularly those of the jump variety, the domain $\mathcal{D}(\mathcal{A})$ is often defined to be a dense subspace of $C_0(\mathbb{R}^d)$, and thus, $\mathcal{D}(\mathcal{A}) \cap \mathcal{F}_i = \{f \equiv 0\}$. Hence, in that case, inequality (3.25) would always trivially hold. Thus, we would like to extend Theorem 3.13 to the extended generator $I(p)$, since $C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$ is non-trivial and dense in $B_b(\mathbb{R}^d) \cap \mathcal{F}_i$.

3.4.3 Small-time asymptotics

The classical results of small-time asymptotics have been primarily established for Lévy processes. For a given Lévy process $L = (L_t)_{t \geq 0}$ it is known that for all $f \in C_c(\mathbb{R}^d \setminus \{0\})$,

$$\lim_{t \searrow 0} \frac{1}{t} \mathbb{E}^0 f(L_t) = \int_{\mathbb{R}^d \setminus \{0\}} f(y) \nu(dy). \tag{3.27}$$

See [28]. Thus, by the Pormanteau theorem, (3.27) implies

$$\lim_{t \searrow 0} \frac{1}{t} \mathbb{P}^0(L_t \in A) = \nu(A)$$

for all $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ with $0 \notin \overline{A}$ and $\nu(\partial A) = 0$. This result naturally extends to a general starting point x: For every $x \in \mathbb{R}^d$,

$$\lim_{t \searrow 0} \frac{1}{t} \mathbb{P}^x(L_t - x \in A) = \nu(A)$$

by translation invariance of a Lévy process. Additional small-time asymptotic results have been given for different classes of functions f in (3.27) (see Sato [44], Jacod [25], Figueroa-López [19]).

Until recently, an analogous statement of the above for Lévy-type processes was not known. However, Kuhn and Schilling (2016) proved in [28] such a statement for Feller processes.

Theorem 3.14 (Kuhn, Schilling [28]). *Let $X = (X_t)_{t\geq 0}$ be a rich Feller process with symbol $p(x,\xi)$ and characteristics $(b(x), Q(x), \nu(x,dy))$. If $f \in C_0(\mathbb{R}^d)$ and $f|_{B(0,\delta)} = 0$ for some $\delta > 0$, then*

$$\lim_{t\searrow 0} \frac{1}{t}\mathbb{E}^x f(X_t - x) = \int_{\mathbb{R}^d\setminus\{0\}} f(y)\nu(x,dy).$$

Additionally, by Pormanteau theorem,

$$\lim_{t\searrow 0} \frac{1}{t}\mathbb{P}^x(X_t - x \in A) = \nu(x,A) \tag{3.28}$$

for all $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ such that $0 \notin \overline{A}$ and $\nu(x,\partial A) = 0$.

The small-time asymptotics given by Theorem 3.14 give us a direct connection between the Lévy measure and the Feller process, surpassing the representation of the generator. Also, notice that the result holds for more general, locally bounded symbols. This will also provide a useful analytical tool in studying the dependence properties in jump-Feller processes.

3.4.4 Main results of this chapter

We aim to prove the following. Consider a rich Feller process $X = (X_t)_{t\geq 0}$ on the space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P}^x)_{x\in\mathbb{R}^d}$ with Lévy characteristics $(b(x), 0, \nu(x,dy))$. If we assume that X is stochastically monotone, then condition

$$\nu(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0, \quad \forall x \in \mathbb{R}^d \tag{3.29}$$

is a necessary and sufficient condition for the association, WA, PSA, PSD, POD, PUOD, and PLOD in space of process X. These equivalences can be illustrated in the implication map in Figure 3.1.

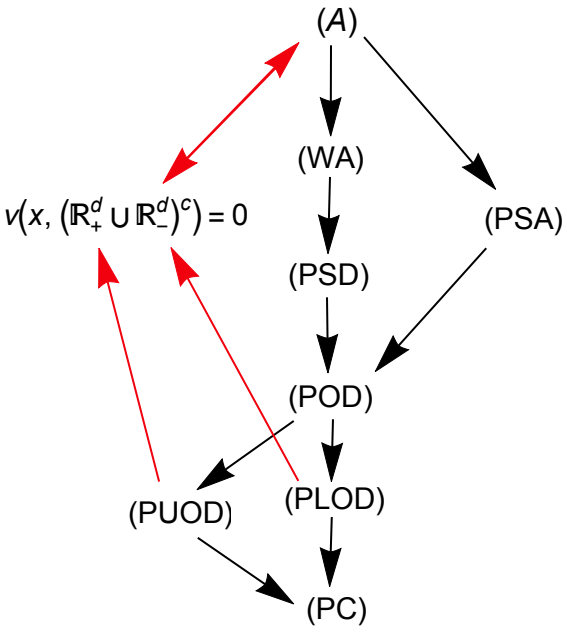

Figure 3.1: Equivalence of dependencies under condition (3.29) for Feller processes

To show these equivalences, we first give a proof that, under stochastic monotonicity, condition (3.29) is equivalent to association in space. We show this in following subsection. Then in the subsequent subsection, we show that PUOD in space implies condition (3.29). We also remark in that subsection that we can also show PLOD in space implies (3.29).

Association equivalent to condition (3.29)

Theorem 3.15 (Tu, 2017a, [50]). *Let $X = (X_t)_{t\geq 0}$ be a stochastically monotone, rich Feller processes with semigroup $(T_t)_{t\geq 0}$, generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$, bounded symbol $p(x,\xi)$, and triplet $(b(x), 0, \nu(x, dy))$. Then*

$$X \text{ is spatially associated with respect to } \mathbb{P}^x, \forall x \in \mathbb{R}^d$$

if and only if

$$\text{condition (3.29): } \forall x,\ \nu(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0, \text{ is satisfied.}$$

We prove this by first showing that spatial association of X is equivalent to a Liggett-type inequality (as given in Theorem 3.13) for the extended generator $I(p)$, the statement of which is in the following theorem.

Theorem 3.16 (Tu, 2017a, [50]). *Let $X = (X_t)_{t\geq 0}$ be a stochastically monotone, rich Feller processes with semigroup $(T_t)_{t\geq 0}$, generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$, bounded symbol $p(x,\xi)$, and an integro-differential operator $I(p)$. Assume $x \mapsto p(x,0)$ is continuous. Then*

$$I(p)fg \geq fI(p)g + gI(p)f, \quad \forall f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i \tag{3.30}$$

if and only if

$$\forall t \geq 0, \quad T_t fg \geq T_t f \cdot T_t g, \quad \forall f, g \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i. \tag{3.31}$$

Inequality (3.31) in Theorem 3.16 is another way to formulate that X is spatially associated, since inequality (3.31) means, for all $x \in \mathbb{R}^d$,

$$\mathbb{E}^x f(X_t)g(X_t) = T_t fg(x) \geq T_t f(x) T_t g(x) = \mathbb{E}^x f(X_t)\mathbb{E}^x g(X_t)$$

which means X_t is associated with respect to $\mathbb{P}^x$. Inequality (3.30) intuitively means that the process either jumps either up or down, which in multidimensional Euclidean space, means that if the process is currently at point x, then it can only jump to another point y if $y \geq x$ or $y \leq x$ componentwise. To see this intuition more clearly, one can write out the inequality for the finite state space case when $I(p)$ would be a Q-(transition) matrix of a Markov chain (See Harris [22]).

Notice that in Theorem 3.16, we are using the extended generator $I(p)$. In previous statements of Liggett's characterization, the generator $\mathcal{A}$ is used. But we need to use $I(p)$ for the reasons given in the comments after Theorem 3.13. Hence, it is necessary to show the Liggett-type inequality as a characterization of association for rich Feller processes. Such an extension has not been seen by the author. We first need the following lemmas to prove Theorem 3.16.

Setting: Let $X = (X_t)_{t\geq 0}$ be a rich Feller process with semigroup $(T_t)_{t\geq 0}$, generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$, symbol $p(x,\xi)$, integro-differential operator $I(p)$, and characteristics $(a(x), b(x), Q(x), \nu(x, dy))$, where b, Q, ν are the same before, except we have an additional

characteristic $a : \mathbb{R}^d \to \mathbb{R}_+$ which represents the "killing rate." This quadruple is actually the more general representation for sub-Markovian Feller processes. See [46], [2].

Remark 3.4.1. With the additional characteristic $a(x)$, the symbol $p(x,\xi)$ would look like

$$p(x,\xi) = -a(x) + ib(x)\cdot\xi - \frac{1}{2}\xi\cdot\Sigma(x)\xi + \int_{y\neq 0}(e^{i\xi\cdot y} - 1 - i\xi\cdot y\chi(y))\nu(x,dy), \tag{3.32}$$

and $I(p)$ would look like

$$\begin{aligned} I(p)f(x) = &-a(x)f(x) + b(x)\cdot\nabla f(x) + \frac{1}{2}\nabla\cdot\Sigma(x)\nabla f(x) \\ &+ \int_{y\neq 0}\left(f(x+y) - f(x) - y\cdot\nabla f(x)\chi(y)\right)\nu(x,dy). \end{aligned} \tag{3.33}$$

When we say $p(x,\xi)$ is bounded, then we also include that $||a(\cdot)||_\infty < \infty$. The function $-p(x,\xi)$ is also a continuous negative definite function [24].

Feller semigroups $(T_t)_{t\geq 0}$ are invariant with respect to $C_0(\mathbb{R}^d)$ and exhibit strong continuity on $C_0(\mathbb{R}^d)$. It is sometimes important however to have semigroups be invariant with respect to $C_b(\mathbb{R}^d)$ and exhibit locally-strong continuity on $C_b(\mathbb{R}^d)$.

Lemma 3.4.1 (Schilling, [45])**.** *Assume the above* ***Setting*** *and additionally that $p(x,\xi)$ is bounded.*

(i) If $x \mapsto p(x,0)$ is continuous, then $T_t 1 \in C_b(\mathbb{R}^d)$.

*(ii) If $T_t 1 \in C_b(\mathbb{R}^d)$, then $(T_t)_{t\geq 0}$ extends to a C_b-**Feller semigroup**, i.e. satisfies*

(a) $T_t : C_b(\mathbb{R}^d) \to C_b(\mathbb{R}^d)$, for all $t \geq 0$,

(b) $\lim_{h\searrow 0} ||T_{t+h}u - T_t u||_{\infty,K} = 0$ for all $K \subset \mathbb{R}^d$ compact, $u \in C_b(\mathbb{R}^d)$, $t \geq 0$, i.e. locally uniformly continuous.

Lemma 3.4.2. *Given the above* ***Setting****,*

$$x \mapsto p(x,0) \text{ is continuous} \qquad \text{if and only if} \qquad I(p) : C_b^2(\mathbb{R}^d) \to C(\mathbb{R}^d). \tag{3.34}$$

Note: $p(x,\xi)$ need not be bounded.

Proof. This notion is stated and proven in Remark 4.5 (ii) of Schilling's paper [45]. We give a more detailed proof here, and in our proof, replace C_b^2 with C_b^∞. We know by Courrege that $p(x,D) = -\mathcal{A}|_{C_c^\infty(\mathbb{R}^d)}$ is a pseudo-differential operator which maps

$$p(x,D) : C_c^\infty(\mathbb{R}^d) \to C(\mathbb{R}^d). \tag{3.35}$$

($\Leftarrow$) Since $1(x) \in C_b^2(\mathbb{R}^d)$, we have that $I(p)1(x) \in C_b(\mathbb{R}^d)$ as a function of x. Observe that

$$\begin{aligned} I(p)1(x) &= -a(x)1(x) + b(x)\cdot\nabla 1(x) + \frac{1}{2}\nabla\cdot Q(x)\nabla 1(x) \\ &\quad + \int_{y\neq 0} \left(1(x+y) - 1(x) - y\cdot\nabla 1(x)\mathbb{1}_{(0,1)}(|y|)\right)\nu(x,dy) \\ &= -a(x) \\ &= -p(x,0). \end{aligned}$$

So $I(p)1(x) \in C_b(\mathbb{R}^d)$ if and only if $x \mapsto p(x,0)$ is continuous.

($\Rightarrow$) Choose a compact set $K \subset \mathbb{R}^d$, a function $f \in C_b^\infty(\mathbb{R}^d)$, and define a sequence $(\phi_k)_{k\in\mathbb{N}}$, such that $\phi \in C_c^\infty(\mathbb{R}^d)$, $0 \le \phi \le 1$, $\phi(0) = 1$, and $\phi = 1$ on $B(0,1)$ and $\phi = 0$ on $B(0,R)^c$, where $1 < R$, and

$$\phi_k(x) := \phi\left(\frac{x}{k}\right),$$

where $\{x : \phi_k(x) = 1\} \nearrow \mathbb{R}^d$ and $\phi_k \nearrow 1$ as $k \to \infty$. Observe that for k sufficiently large, we would have $K \subset \{x : \phi_k(x) = 1\}$, which means $\phi_k = 1$ on K. Thus, for sufficiently large k, we have

$$\begin{aligned} &\mathbb{1}_K(x)\,|I(p)f(x) - I(p)(\phi_k f)(x)| \\ &= \mathbb{1}_K(x)\left|-a(x)(f(x)-\phi_k f(x)) + b(x)\cdot\nabla(f(x)-\phi_k(x)f(x)) + \frac{1}{2}\nabla\cdot Q(x)\nabla(f(x)-\phi_k(x)f(x)\right. \\ &\quad \left. + \int_{y\neq 0}\left((f(x+y)-\phi_k f(x+y)) - (f(x)-\phi_k f(x)) - y\cdot\nabla(f(x)-\phi_k f(x))\mathbb{1}_{(0,1)}(|y|)\right)\nu(x,dy)\right| \\ &= \mathbb{1}_K(x)\left|\int_{y\neq 0}(f(x+y)-\phi_k(x+y)f(x+y))\nu(x,dy)\right| \\ &= \mathbb{1}_K(x)\left|\int_{y\neq 0} f(x+y)(1-\phi_k(x+y))\nu(x,dy)\right| \end{aligned}$$

$$\leq \mathbb{1}_K(x)||f||_\infty \int_{y\neq 0}(1-\phi_k(x+y))\nu(x,dy)$$

where we get the first equality because $\phi_k = 1$ on this set (sufficiently large k). Now to show this integral goes to 0, we need a dominated convergence argument. First observe that by change of variables, we have

$$\begin{aligned}\int_{y\neq 0}(1-\phi_k(x+y))\nu(x,dy) &= \int_{z\neq 0}(1-\phi_k(x-z))\nu(x,-dz)\\ &= \int_{z\neq 0}(1-\phi_k(x-z))\hat{\nu}(x,dz)\end{aligned}$$

where $\hat{\nu}(x,A) := \nu(x,-A)$, and $\hat{\nu}(x,\cdot)$ is still a σ-finite measure with

$$\int_{z\neq 0} 1\wedge|z|^2\hat{\nu}(x,dz) = \int_{z\neq 0} 1\wedge|z|^2\nu(x,-dz) = \int_{y\neq 0} 1\wedge|y|^2\nu(x,dy) < \infty.$$

Observe that the integrand of $\int_{z\neq 0}(1-\phi_k(x-z))\hat{\nu}(x,dz)$ vanishes on $y\in B(x,k)$ or $|y-x|<k$. This is because $|y-x|<k$ implies $\left|\dfrac{x-y}{k}\right|<1$, which means $\phi_k(x-y)=\phi\left(\dfrac{x-y}{k}\right)=1$.

Now for sufficiently large k, $B(x,k)\supset B(0,1)$, thus

$$\begin{aligned}\int_{z\neq 0}(1-\phi_k(x-z))\hat{\nu}(x,dz) &= \int_{B(x,k)^c}(1-\phi_k(x-z))\hat{\nu}(x,dz)\\ &\leq \int_{B(x,k)^c} 1\hat{\nu}(x,dz)\\ &\leq \int_{B(0,1)^c} 1\hat{\nu}(x,dz)\\ &= \int_{|z|\geq 1} 1\hat{\nu}(x,dz)\\ &= \int_{z\neq 0} 1\wedge|z|^2\hat{\nu}(x,dz) < \infty\end{aligned}$$

by assumption. Therefore, by dominated convergence theorem,

$$\mathbb{1}_K(x)||f||_\infty \int_{y\neq 0}(1-\phi_k(x+y))\nu(x,dy)$$

$$= \mathbb{1}_K(x)||f||_\infty \int_{y\neq 0} (1-\phi_k(x-y))\hat{\nu}(x,dy)$$
$$\to 0$$

Thus, we have convergence

$$I(p)\phi_k f \to I(p)f$$

locally uniformly. Now note that if $\phi_k \in C_c^\infty(\mathbb{R}^d)$ and $f \in C_b^\infty(\mathbb{R}^d)$, then $\phi_k f \in C_c^\infty(\mathbb{R}^d)$ for all k. By (3.35), we have that $\{I(p)\phi_k f\}_{k\in\mathbb{N}} = \{p(x,D)\phi_k f\}_{k\in\mathbb{N}} \subset C(\mathbb{R}^d)$. Since continuity is preserved under locally uniform convergence, we have that $I(p)f \in C(\mathbb{R}^d)$. □

Corollary 3.4.2. *Given the above setting, if $p(x,\xi)$ is bounded, then $x \mapsto p(x,0)$ is continuous if and only if $I(p) : C_b^2(\mathbb{R}^d) \to C_b(\mathbb{R}^d)$.*

Proof. All that is needed to show is if $x \mapsto p(x,0)$ is continuous, then $I(p)u$ is a bounded function if $u \in C_b^2(\mathbb{R}^d)$. This can easily be shown using the representation in (3.23) and is proven in [45]. □

Lemma 3.4.3. *Assume the above **Setting** and that $p(x,\xi)$ is bounded. If $x \mapsto p(x,0)$ is continuous, then $I(p)$ generates the semigroup $(T_t)_{t\geq 0}$ locally uniformly, i.e.*

$$I(p)f = \lim_{t\searrow 0} \frac{1}{t}(T_t f - f), \qquad f \in C_b^2(\mathbb{R}^d) \tag{3.36}$$

where convergence is locally uniform.

Proof. By a result of Schilling [46], the process

$$M_t^f := f(X_t) - f(X_0) - \int_0^t I(p)f(X_{s-})ds \tag{3.37}$$

is, for every $f \in C_b^2(\mathbb{R}^d)$, an L^2-local martingale with respect to $\mathbb{P}^x$, for all x. This implies

$$0 = \mathbb{E}^x M_t^f$$
$$= \mathbb{E}^x f(X_t) - \mathbb{E}^x f(X_0) - \mathbb{E}^x \int_0^t I(p)f(X_{s-})ds$$
$$= T_t f(x) - f(x) - \int_0^t \mathbb{E}^x I(p)f(X_{s-})ds$$

$$= T_t f(x) - f(x) - \int_0^t T_s I(p) f(x) ds$$

for every $x \in \mathbb{R}^d$, $t \geq 0$. Note that we can switch integrals in line 3 because $I(p)f \in C_b(\mathbb{R}^d)$ by Lemma 3.4.2 and Corollary 3.4.2 This implies

$$\frac{1}{t}(T_t f - f) = \frac{1}{t}\int_0^t T_s I(p) f \, ds. \tag{3.38}$$

We argue that when taking the limit as $t \searrow 0$, the right hand-side converges locally uniformly to $I(p)f$. Note that since $I(p)f \in C_b(R^d)$, then $(T_s I(p)f)\mathbb{1}_K$ is continuous in s for every compact set K by the C_b-Feller property, i.e.

$$\begin{aligned} ||(T_{s+h} I(p)f)\mathbb{1}_K - (T_s I(p)f)\mathbb{1}_K||_\infty &= ||(T_{s+h} I(p)f - T_s I(p)f)\mathbb{1}_K||_\infty \\ &= \sup_{x \in K} |T_{s+h} I(p)f(x) - T_s I(p)f(x)| \\ &\to 0 \end{aligned}$$

So, the function $T_{(\cdot)} I(p) f \mathbb{1}_K$ is the integrand of a Bochner-type integral (see [15]) that is continuous in s and integrable on any closed interval $[a, b]$. Therefore, by Fundamental Theorem of Calculus for Bochner integrals,

$$\begin{aligned} \lim_{t \searrow 0} \frac{1}{t}(T_t f - f)\mathbb{1}_K &= \lim_{t \searrow 0} \frac{1}{t} \int_0^t (T_s I(p) f)\mathbb{1}_K \, ds \\ &= (I(p)f)\mathbb{1}_K \end{aligned}$$

for all $K \subset \mathbb{R}^d$ compact. Hence, $I(p)f = \lim_{t \searrow 0} \frac{1}{t}(T_t f - f)$, where convergence is locally uniform. □

Remark 3.4.2. The author has also proved, using similar techniques and additionally the technique of "stopped processes" that $I(p)$ generates $(T_t)_{t \geq 0}$ pointwise, i.e. where the convergence in (3.36) is pointwise. Also, the above lemma is proven when $p(x, \xi)$ is just

locally bounded, but also satisfying

$$\lim_{k\to\infty} \sup_{|y-x|\le 2k} \sup_{|\eta|\le 1/k} |p(y,\eta)| = 0, \qquad \forall x \in \mathbb{R}^d \tag{3.39}$$

Lemma 3.4.4. *Assume the above **Setting**, and the symbol $p(x,\xi)$ is bounded and $x \mapsto p(x,0)$ is continuous. For all $f \in C_b^2(\mathbb{R}^d)$,*

$$\frac{d}{dt}T_t f = I(p)T_t f = T_t I(p) f$$

where the derivative is defined based on locally uniform convergence.

Proof. Choose $f \in C_b^2(\mathbb{R}^d)$. Then for all $t \ge 0$, $T_t f \in C_b^2(\mathbb{R}^d)$ by the C_b-Feller property. Hence,

$$\lim_{h\to 0} \frac{T_{t+h}f - T_t f}{h} = \lim_{h\to 0} \frac{1}{h}(T_h(T_t f) - T_t f) = I(p)T_t f$$

by Lemma 3.4.3, where convergence is locally uniform. Now observe that for all $x \in \mathbb{R}^d$,

$$\begin{aligned}
T_{t+h}f(x) - T_t f(x) &= T_t(T_h f(x) - f(x)) \\
&= T_t \int_0^h T_s I(p) f(x)\, ds \\
&= \mathbb{E}^x \int_0^h T_s I(p) f(X_t)\, ds \\
&= \int_0^h \mathbb{E}^x T_s I(p) f(X_t)\, ds, \text{ by Fubini's Theorem,} \\
&= \int_0^h T_t T_s I(p) f(x)\, ds \\
&= \int_0^h T_s T_t I(p) f(x)\, ds.
\end{aligned}$$

Thus,

$$\lim_{h\to 0} \frac{1}{h}(T_{t+h}f - T_t f) = \lim_{h\to 0} \frac{1}{h} \int_0^h T_s T_t I(p) f\, ds$$

$$= T_t I(p) f$$

because $T_t I(p) f \in C_b(\mathbb{R}^d)$ by C_b-Feller property, thus making $T_s T_t I(p) f \mathbb{1}_K$ continuous in s for every compact K. Once again, by Fundamental Theorem of Calculus for Bochner integrals, we get the convergence shown above. As usual, the limits taken above are with respect to locally uniform convergence.

□

Finally, in the proof of Liggett's Theorem (Theorem 3.13), Liggett makes use of a solution to a Cauchy problem, which he proves using a bounded generator that generates a semigroup uniformly. We would like to extend this result to the integro-differential operator $I(p)$ that generates the semigroup locally uniformly. This is given below.

Lemma 3.4.5 (Extension of Liggett's Cauchy problem)**.** *Let $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ be a (rich) Feller generator of a semigroup $(T_t)_{t\geq 0}$ with bounded Lévy coefficients and symbol $p(x,\xi)$ satisfying $x \mapsto p(x,0)$ continuous (these assumptions give us C_b-Feller property). Let $I(p)$ be the extended generator on $C_b^2(\mathbb{R}^d)$. Suppose $F, G : [0,\infty) \to C_b(\mathbb{R}^d)$ such that*
(a) $F(t) \in \mathcal{D}(I(p))$ for all $t \geq 0$,
(b) $G(t)$ is continuous on $[0,\infty)$ (locally uniformly),
(c) $F'(t) = I(p)F(t) + G(t)$ for all $t \geq 0$.

Then

$$F(t) = T_t F(0) + \int_0^t T_{t-s} G(s) ds.$$

Proof. Observe that all limits (and corresponding derivatives) we take here are with respect to locally uniform convergence. Also, by Lemma 3.4.3, we have

$$\lim_{t \searrow 0} \frac{1}{t}(T_t u - u) = I(p)u$$

for all $u \in C_b^2(\mathbb{R}^d)$. Also, observe that we will define the derivative $F'(s)$ by

$$F'(s) = \lim_{h \to 0} \frac{F(s+h) - F(s)}{h}$$

where the limit is under locally uniform convergence. Also, our statement of (b) is different then Liggett's.

$$\text{Liggett's: if } t_n \to t, \text{ then } ||G(t_n) - G(t)||_\infty \to 0 \text{ as } n \to \infty.$$

$$\text{Ours: if } t_n \to t, \text{ then } ||G(t_n) - G(t)||_{\infty,K} \to 0 \text{ as } n \to \infty \text{ for all } K \text{ compact.}$$

Though Liggett's assumption would be sufficient, we don't need something that strong in our setting, and our G will satisfy the locally uniform continuity. Choose some compact set $K \subset \mathbb{R}^d$.

$$\begin{aligned}
&\frac{T_{t-s-h}F(s+h) - T_{t-s}F(s)}{h} \cdot \mathbb{1}_K \\
&= \frac{T_{t-s-h}F(s+h)}{h} \cdot \mathbb{1}_K - \frac{T_{t-s}F(s)}{h} \cdot \mathbb{1}_K \\
&+ [T_{t-s-h} - T_{t-s}]F'(s) \cdot \mathbb{1}_K - [T_{t-s-h} - T_{t-s}]F'(s) \cdot \mathbb{1}_K \\
&+ \frac{T_{t-s-h}F(s)}{h} \cdot \mathbb{1}_K - \frac{T_{t-s-h}F(s)}{h} \cdot \mathbb{1}_K \\
&+ \frac{T_{t-s}F(s+h)}{h} \cdot \mathbb{1}_K - \frac{T_{t-s}F(s+h)}{h} \cdot \mathbb{1}_K \\
&+ \frac{T_{t-s}F(s)}{h} \cdot \mathbb{1}_K - \frac{T_{t-s}F(s)}{h} \cdot \mathbb{1}_K \\
&=: (1) + (2) + (3) + (4) + (5) + (6) + (7) + (8) + (9) + (10) \\
&= [(2) + (7)] + [(5) + (10)] + (3) + [(4) + (1) + (9) + (8) + (6)] \\
&= T_{t-s}\left[\frac{F(s+h) - F(s)}{h}\right] \cdot \mathbb{1}_K + \left[\frac{T_{t-s-h} - T_{t-s}}{h}\right] F(s) \cdot \mathbb{1}_K \\
&+ [T_{t-s-h} - T_{t-s}]F'(s) \cdot \mathbb{1}_K + [T_{t-s-h} - T_{t-s}]\left[\frac{F(s+h) - F(s)}{h} - F'(s)\right] \cdot \mathbb{1}_K \\
&=: (I) + (II) + (III) + (IV).
\end{aligned}$$

Now we consider the limits as h goes to 0 for each of these four terms.

(I):

$$\lim_{h \searrow 0} T_{t-s}\left[\frac{F(s+h) - F(s)}{h}\right] \cdot \mathbb{1}_K = T_{t-s} \lim_{h \searrow 0}\left[\frac{F(s+h) - F(s)}{h}\right] \cdot \mathbb{1}_K$$

$$= T_{t-s}F'(s) \cdot \mathbb{1}_K$$

because T_{t-s} is a bounded operator, which means it is a continuous operator.

(II): Let $u = t - s$. Then $s = t - u$ and $ds = -du$. For a function $f \in C_b(\mathbb{R}^d)$,

$$\begin{aligned}\lim_{h \searrow 0} \left[\frac{T_{t-s-h} - T_{t-s}}{h}\right] f \cdot \mathbb{1}_K &= \frac{d}{ds} T_{t-s} f \cdot \mathbb{1}_K \\ &= -\frac{d}{du} T_u f \cdot \mathbb{1}_K \\ &= -I(p) T_u f \cdot \mathbb{1}_K \\ &= -I(p) T_{t-s} f \cdot \mathbb{1}_K.\end{aligned}$$

Therefore,

$$\lim_{h \searrow 0} \left[\frac{T_{t-s-h} - T_{t-s}}{h}\right] F(s) \cdot \mathbb{1}_K = -I(p) T_{t-s} F(s) \cdot \mathbb{1}_K = -T_{t-s} I(p) F(s) \cdot \mathbb{1}_K.$$

(III): By C_b-Feller property, since $F'(s) \in C_b(\mathbb{R}^d)$ (in our setting),

$$\lim_{h \searrow 0} [T_{t-s-h} - T_{t-s}] F'(s) \cdot \mathbb{1}_K = 0$$

uniformly.

(IV): Observe that T_{t-s-h} and T_{t-s} are both contractions. Hence,

$$\begin{aligned}&\left|\left|[T_{t-s-h} - T_{t-s}] \left[\frac{F(s+h) - F(s)}{h} - F'(s)\right]\right|\right|_{\infty, K} \\ &\leq ||T_{t-s-h} - T_{t-s}|| \cdot \left|\left|\left[\frac{F(s+h) - F(s)}{h} - F'(s)\right]\right|\right|_{\infty, K} \\ &\leq 2 \left|\left|\left[\frac{F(s+h) - F(s)}{h} - F'(s)\right]\right|\right|_{\infty, K} \\ &\to 0\end{aligned}$$

as $h \to 0$. Thus, we have for $0 < s < t$,

$$\begin{aligned}
\frac{d}{ds}T_{t-s}F(s)\cdot\mathbb{1}_K &= \lim_{h\searrow 0}\frac{T_{t-(s+h)}F(s+h)-T_{t-s}F(s)}{h}\cdot\mathbb{1}_K \\
&= \lim_{h\searrow 0}[(I)+(II)+(III)+(IV)] \\
&= T_{t-s}F'(s)\cdot\mathbb{1}_K - T_{t-s}I(p)F(s)\cdot\mathbb{1}_K \\
&= T_{t-s}[F'(s)-I(p)F(s)]\cdot\mathbb{1}_K \\
&\overset{(c)}{=} T_{t-s}G(s)\cdot\mathbb{1}_K.
\end{aligned}$$

The right-hand side is a continuous function of s because G is continuous function of s and the semigroup is uniformly continuous on K by C_b-Feller property. Let's justify this:

Aside: Let $\epsilon > 0$. Then $\exists N$ large s.t. $||G(s_n)-G(s)||_{\infty,K} < \epsilon/2$ for all $n \geq N$. Also, $\exists N'$ large s.t.

$$||T_{t-s_n}G(s_N)-T_{t-s}G(s_N)||_{\infty,K} = ||(T_{t-s_n}-T_{t-s})G(s_N)||_{\infty,K} < \epsilon/2$$

for all $n \geq N'$ since semigroup operator is uniformly continuous on compact sets. Let $M = \max(N, N')$. Then

$$\begin{aligned}
||T_{t-s_M}G(s_M)-T_{t-s}G(s)||_{\infty,K} &= ||T_{t-s_M}G(s_M)-T_{t-s}G(s_M)+T_{t-s}G(s_M)-T_{t-s}G(s)||_{\infty,K} \\
&\leq ||T_{t-s_M}G(s_M)-T_{t-s}G(s_M)||_{\infty,K} + ||T_{t-s}G(s_M)-T_{t-s}G(s)||_{\infty,K} \\
&\leq ||T_{t-s_M}G(s_M)-T_{t-s}G(s_M)||_{\infty,K} + ||G(s_M)-G(s)||_{\infty,K} \\
&< \epsilon/2+\epsilon/2 = \epsilon.
\end{aligned}$$

Therefore we can integrate these functions with respect to s from 0 to t. And by Fundamental Theorem of Calculus,

$$\begin{aligned}
\int_0^t \frac{d}{ds}T_{t-s}F(s)ds\cdot\mathbb{1}_K &= \int_0^t T_{t-s}G(s)ds\cdot\mathbb{1}_K \\
(T_{t-t}F(t)-T_tF(0))\cdot\mathbb{1}_K &= \int_0^t T_{t-s}G(s)ds\cdot\mathbb{1}_K
\end{aligned}$$

$$F(t) \cdot \mathbb{1}_K = \left(T_t F(0) + \int_0^t T_{t-s} G(s) ds \right) \cdot \mathbb{1}_K.$$

Since K compact arbitrary, we have our desired result $F(t) = T_t F(0) + \int_0^t T_{t-s} G(s) ds$ for all $t \geq 0$. □

We are now ready to prove the main theorems of this section.

Proof of Theorem 3.16

Proof. ($\Leftarrow$) Assume $T_t fg \geq T_t f \, T_t g$ for all $f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$. This implies

$$\begin{aligned} T_t fg - fg &\geq T_t f \, T_t g - fg \\ &= T_t f \, T_t g - fg + g \, T_t f - g \, T_t f \\ &= T_t f [T_t g - g] + g[T_t f - f]. \end{aligned}$$

Hence, for all $t > 0$,

$$\frac{1}{t}(T_t fg - fg) \geq T_t f \, \frac{T_t g - g}{t} + g \, \frac{T_t f - f}{t}.$$

Therefore,

$$\begin{aligned} I(p)fg &= \lim_{t \searrow 0} \frac{1}{t}(T_t fg - fg) \\ &\geq \lim_{t \searrow 0} \left(T_t f \, \frac{T_t g - g}{t} + g \, \frac{T_t f - f}{t} \right) \\ &= \left(\lim_{t \searrow 0} T_t f \right) \left(\lim_{t \searrow 0} \frac{T_t g - g}{t} \right) + g \left(\lim_{t \searrow 0} \frac{T_t f - f}{t} \right) \\ &= f I(p) g + g I(p) f \end{aligned}$$

where the convergence is locally uniform.

($\Rightarrow$) Assume $I(p)fg \geq f I(p) g + g I(p) f$ for all $f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$. By monotonicity, $T_t f, T_t g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$, which implies

$$I(p)(T_t f)(T_t g) \geq T_t f [I(p) T_t g] + T_t g [I(p) T_t f]. \tag{3.40}$$

Define $F(t) = T_t fg - T_t f\, T_t g$. Then by Lemma 3.4.4, we have

$$\begin{aligned} F'(t) &= \frac{d}{dt} T_t fg - \frac{d}{dt} T_t f\, T_t g \\ &= I(p) T_t fg - (T_t f [I(p) T_t g] + T_t g [I(p) T_t f]) \\ &\geq I(p) T_t fg - (I(p) T_t f\, T_t g) \\ &= I(p)(T_t fg - T_t f\, T_t g) \\ &= I(p) F(t), \end{aligned}$$

where the inequality comes from (3.40). Define $G(t) := F'(t) - I(p)F(t) \geq 0$. Then by Lemma 3.4.5, the solution to the Cauchy problem $F'(t) = G(t) + I(p)F(t)$ is given by

$$F(t) = T_t F(0) + \int_0^t T_{t-s} G(s) ds = \int_0^t T_{t-s} G(s) ds$$

since $F(0) = 0$. Since $G(s) \geq 0$ for all s, and T_{t-s} is a positivity-preserving linear operator, $F(t) \geq 0$ for all $t \geq 0$. Thus, $T_t fg \geq T_t f \cdot T_t g$ for all $f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$. This inequality also holds for all $f, g \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i$, since we can approximate non-decreasing, continuous, bounded functions by non-decreasing smooth, bounded functions, and then use a dominated convergence argument (see proof of Proposition 2.1.1 (c)). □

Remark 3.4.3.

1. Note that in the statement of Theorem 3.16, we assumed $x \mapsto p(x, 0)$ must be continuous. But it we assume a Feller process (not sub-Markovian) with $a(x) = 0$, then we are already guaranteed this.

2. For the necessary condition, we did not need stochastic monotonicity.

Proof of Theorem 3.15

Proof. ($\Leftarrow$). Fix $x \in \mathbb{R}^d$. Assume $\nu(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0$. Then, for all $f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$,

$$\begin{aligned} &I(p)fg(x) - g(x)I(p)f(x) - f(x)I(p)g(x) \\ &= b(x) \cdot \nabla fg(x) + \int_{y \neq 0} \left(f(x+y)g(x+y) - f(x)g(x) - y \cdot \nabla fg(x) \mathbb{1}_{(0,1)}(|y|) \right) \nu(x, dy) \end{aligned}$$

$$
\begin{aligned}
&- b(x)\cdot g(x)\nabla f(x) - \int_{y\neq 0}\left(f(x+y)g(x) - f(x)g(x) - y\cdot g(x)\nabla f(x)\mathbb{1}_{(0,1)}(|y|)\right)\nu(x,dy)\\
&- b(x)\cdot f(x)\nabla g(x) - \int_{y\neq 0}\left(f(x)g(x+y) - f(x)g(x) - y\cdot f(x)\nabla g(x)\mathbb{1}_{(0,1)}(|y|)\right)\nu(x,dy)\\
&= \int_{y\neq 0}\left(f(x+y)g(x+y) - f(x+y)g(x) - f(x)g(x+y) + f(x)g(x)\right)\nu(x,dy)\\
&= \int_{y\neq 0}\left(f(x+y) - f(x)\right)\left(g(x+y) - g(x)\right)\nu(x,dy)\\
&= \int_{\mathbb{R}^d_+}\left(f(x+y) - f(x)\right)\left(g(x+y) - g(x)\right)\nu(x,dy)\\
&\quad + \int_{\mathbb{R}^d_-}\left(f(x+y) - f(x)\right)\left(g(x+y) - g(x)\right)\nu(x,dy)\\
&\geq 0,
\end{aligned}
$$

where the drift terms and the cut-off term in the integrand vanish because $\nabla fg(x) = f(x)\nabla g(x) + g(x)\nabla f(x)$. Additionally, we get positivity at the end there because $\forall y \in \mathbb{R}^d_+$, $f(x+y) - f(x) \geq 0$, and $g(x+y) - g(x) \geq 0$, so $(f(x+y) - f(x))(g(x+y) - g(x)) \geq 0$ on $\mathbb{R}^d_+$. A similar result holds on $\mathbb{R}^d_-$. By Theorem 3.16, this implies $T_t fg(x) \geq T_t f(x) T_t g(x)$, where $f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$.

Now to obtain association of X_t, this inequality needs to hold for all $f, g \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i$ But we can use an approximation of a function $f \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i$ by $f_n \in C_b^\infty(\mathbb{R}^d) \cap \mathcal{F}_i$ which gives us the desired result.

($\Rightarrow$). Assume X_t is associated for all $t \geq 0$. This means $T_t fg(x) \geq T_t f(x) T_t g(x)$ for all $x \in \mathbb{R}^d$, for all $f, g \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i$. So this inequality of course holds for $f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$, which yields $I(p)fg \geq gI(p)f + fI(p)g$ for such f, g by Theorem 3.16. This implies, by a similar calculation in the ($\Leftarrow$) direction, that

$$
\int_{y\neq 0}\left(f(x+y) - f(x)\right)\left(g(x+y) - g(x)\right)\nu(x,dy) \geq 0.
$$

For simplicity, assume $d = 2$, but know that we can easily generalize this result to higher dimensions using correction functions. Fix $x = (x_1, x_2) \in \mathbb{R}^2$. Assume for contradiction that Resnick's condition is not satisfied. WLOG, let's say $\nu(x, (0,\infty) \times (-\infty, 0)) > 0$. By

continuity of measure, $\exists a > 0$ such that $\nu(x, (a, \infty) \times (-\infty, a)) > 0$. Let $\epsilon \in (0, 1)$, and define $f, g \in C_b^\infty(\mathbb{R}^2) \cap \mathcal{F}_i$ by

$$f(y_1, y_2) = \begin{cases} 0 & \text{if} \quad y_1 \leq x_1 + \epsilon a \\ 1 & \text{if} \quad y_1 \geq x_1 + a \end{cases}$$

$$g(y_1, y_2) = \begin{cases} 0 & \text{if} \quad y_2 \geq x_2 - \epsilon a \\ -1 & \text{if} \quad y_2 \leq x_2 - a \end{cases}$$

This implies $f(x) = g(x) = 0$. Hence,

$$\begin{aligned}
0 &\leq \int_{y \neq 0} (f(x+y) - f(x))(g(x+y) - g(x))\nu(x, dy) \\
&= \int_{y \neq 0} f(x+y)g(x+y)\nu(x, dy) \\
&= \int_{(a,\infty)\times(-\infty,-a)} f(x+y)g(x+y)\nu(x, dy) + \int_{(a,\infty)\times[-a,-\epsilon a]} f(x+y)g(x+y)\nu(x, dy) \\
&\quad + \int_{[\epsilon a,a]\times(-\infty,-a)} f(x+y)g(x+y)\nu(x, dy) + \int_{[\epsilon a,a]\times[-a,-\epsilon a]} f(x+y)g(x+y)\nu(x, dy) \\
&= -\nu(x, (a, \infty) \times (-\infty, -a)) - \int_{(a,\infty)\times[-a,-\epsilon a]} g(x+y)\nu(x, dy) \\
&\quad + \int_{[\epsilon a,a]\times(-\infty,-a)} f(x+y)\nu(x, dy) + \int_{[\epsilon a,a]\times[-a,-\epsilon a]} f(x+y)g(x+y)\nu(x, dy) \\
&\leq -\nu(x, (a, \infty) \times (-\infty, -a))
\end{aligned}$$

which implies $\nu(x, (a, \infty) \times (-\infty, -a)) \leq 0$, which means $\nu(x, (a, \infty) \times (-\infty, -a)) = 0$, a contradiction. □

Corollary 3.4.3. *Let $X = (X_t)_{t \geq 0}$ be a stochastically monotone, rich Feller process with bounded symbol $p(x, \xi)$, characteristics $(b(x), 0, \nu(x, dy))$. Assume $X_0 \sim \mu \in \mathcal{M}_a$. If either one of the equivalent conditions (3.29) or spatial association holds, then X is also associated in time.*

Proof. True by Theorem 3.15 and Theorem 3.7. □

These proofs complete the notion that spatial association is equivalent to condition (3.29) for jump-Feller processes. Now we want to show condition (3.29) is also necessary for spatial PUOD (and also for spatial PLOD).

PUOD implies condition (3.29)

Lemma 3.4.6. *If $Y = (Y_1, ..., Y_d)$ is POD (PUOD, PLOD), then $(Y_{k_1}, ..., Y_{k_n})$ is POD (PUOD, PLOD, respectively), for all multi-indices $\{k_j\}_{j=1}^n \subset \{1, ..., d\}$.*

Proof. We just show the proof for PUOD. If Y PUOD, then we know $\mathbb{E}\left(\prod_{i=1}^d f_i(Y_i)\right) \geq \prod_{i=1}^d \mathbb{E}f_i(Y_i)$ where $f_i : \mathbb{R} \to \mathbb{R}_+$ non-decreasing. So for all $i \in \{1, ..., d\} \setminus \{k_j\}_{j=1}^n$, set $f_i = \mathbb{1}_{\mathbb{R}}$. Then the above inequality becomes

$$\mathbb{E}\left(\prod_{j=1}^n f_j(Y_{k_j})\right) \geq \prod_{j=1}^n \mathbb{E}f_j(Y_{k_j}).$$

Thus, we have that $(Y_{k_1}, ..., Y_{k_n})$ is PUOD. □

Lemma 3.4.7. *Let ν be a σ-finite measure. Then the set $\{r > 0 : \nu(\partial B(0, r)) = 0\}$ is uncountable infinite.*

Proof. We begin with the easy case that ν is a probability measure. Call $A_r = \partial B(0, r)$. Then we know that for all n, $\text{Card}\{r > 0 : \nu(A_r) > 1/n\} \leq n$, because if it wasn't then the cardinality would be greater than or equal to $n + 1$. But observe that

$$\begin{aligned} 1 &\geq \nu(A_{r_1} \cup ... \cup A_{r_{n+1}}) \\ &= \sum_{k=1}^{n+1} \nu(A_{r_k}), \text{ by disjointness} \\ &> \frac{n+1}{n} \\ &> 1 \end{aligned}$$

a contradiction. Thus,

$$\text{Card}\{r > 0 : \nu(A_r) > 0\} = \text{Card}\left(\bigcup_{n \in \mathbb{N}} \{r > 0 : \nu(A_r) > 1/n\}\right)$$

which must be countable, since it is the union of a countable collection of countable sets. Hence, Card$\{r > 0 : \nu(\partial B(0,r)) = 0\}$ is uncountably infinite. Next, we can easily see how to extend this to finite measures ν. Then we could do this on σ-finite measures by breaking up the space into a countable disjoint union of sets with finite measure. Then, we can carry our the proof on each of the finite-measure sets, then consider their countable union, yielding the same result. Observe that this proof was done for the boundary of open balls. But we can generalize to boundary of any Borel set. □

Theorem 3.17 (Tu, 2017a, [50])**.** *Let $X = (X_t)_{t\geq 0}$ be a rich Feller process with symbol $p(x,\xi)$ and triplet $(b(x), 0, \nu(x,dy))$. Then X_t is PUOD for each $t \geq 0$ implies condition (3.29): $\nu(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0$.*

Proof. Assume X_t is PUOD (wrt $\mathbb{P}^x$) for each $t \geq 0$. Fix $x = (x_1, ..., x_d) \in \mathbb{R}^d$. Since X_t is PUOD, then $X_t - x$ is PUOD for all $t \geq 0$. Assume for contradiction that ν not concentrated on $\mathbb{R}^d_+ \cup \mathbb{R}^d_-$. WLOG, say $\nu(x, (0,\infty)^{d-1} \times (-\infty, 0)) > 0$. By continuity of measure and Lemma 3.4.7, there exists $a > 0$ such that

$$\nu(x, (a,\infty)^{d-1} \times (-\infty, -a)) > 0$$

and

$$\nu(x, \partial[(a,\infty)^{d-1} \times (-\infty, -a)]) = \nu(x, \partial[(a,\infty) \times \mathbb{R}^{d-1}]) = 0.$$

Then by Theorem 3.14,

$$\lim_{t\to 0} \frac{1}{t} \mathbb{P}^x(X_t - x \in (a,\infty)^{d-1} \times (-\infty, -a)) = \nu(x, (a,\infty)^{d-1} \times (-\infty, -a)).$$

$$\begin{aligned}
0 &< \nu(x, (a,\infty)^{d-1} \times (-\infty, -a)) \\
&= \lim_{t\to 0} \frac{1}{t} \mathbb{P}^x(X_t - x \in (a,\infty)^{d-1} \times (-\infty, -a)) \\
&= \lim_{t\to 0} \frac{1}{t} \mathbb{P}^x(X_t^{(1)} - x_1 > a, ..., X_t^{(d-1)} - x_{d-1} > a, X_t^{(d)} - x_d < -a) \\
&\leq \lim_{t\to 0} \frac{1}{t} \mathbb{P}^x(X_t^{(1)} - x_1 > a, ..., X_t^{(d-1)} - x_{d-1} > a, X_t^{(d)} - x_d \leq -a) \\
&= \lim_{t\to 0} \frac{1}{t} \mathbb{P}^x(\{X_t^{(1)} - x_1 > a\} \setminus [\{X_t^{(1)} - x_1 > a\} \cap \{X_t^{(2)} - x_2 > a, ..., X_t^{(d)} - x_d \leq -a\}^c])
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{t\to 0}\frac{1}{t}\left[\mathbb{P}^x(X_t^{(1)}-x_1>a)-\mathbb{P}^x(\{X_t^{(1)}-x_1>a\}\cap\{X_t^{(2)}-x_2>a,...,X_t^{(d)}-x_d\leq -a\}^c)\right]\\
&= \lim_{t\to 0}\frac{1}{t}\Big[\mathbb{P}^x(X_t^{(1)}-x_1>a)\\
&\qquad -\mathbb{P}^x(\{X_t^{(1)}-x_1>a\}\cap[\{X_t^{(2)}-x_2\leq a\}\cup...\cup\{X_t^{(d)}-x_d>-a\}])\Big]\\
&= \lim_{t\to 0}\frac{1}{t}\Big[\mathbb{P}^x(X_t^{(1)}-x_1>a)\\
&\qquad -\mathbb{P}^x(\{X_t^{(1)}-x_1>a,X_t^{(2)}-x_2\leq a\}\cup...\cup\{X_t^{(1)}-x_1>a,X_t^{(d)}-x_d>-a\})\Big]\\
&\leq \lim_{t\to 0}\frac{1}{t}\left[\mathbb{P}^x(X_t^{(1)}-x_1>a)-\mathbb{P}^x(X_t^{(1)}-x_1>a,X_t^{(d)}-x_d>-a)\right]\\
&\leq \lim_{t\to 0}\frac{1}{t}\left[\mathbb{P}^x(X_t^{(1)}-x_1>a)-\mathbb{P}^x(X_t^{(1)}-x_1>a)\mathbb{P}^x(X_t^{(d)}-x_d>-a])\right]\\
&= \lim_{t\to 0}\frac{1}{t}\left[\mathbb{P}^x(X_t^{(1)}-x_1>a)(1-\mathbb{P}^x(X_t^{(d)}-x_d>-a))\right]\\
&= \lim_{t\to 0}\frac{1}{t}\mathbb{P}^x(X_t^{(1)}-x_1>a)\mathbb{P}^x(X_t^{(d)}-x_d\leq -a)\\
&= \left[\lim_{t\to 0}\frac{1}{t}\mathbb{P}^x(X_t^{(1)}-x_1>a)\right]\left[\lim_{t\to 0}\mathbb{P}^x(X_t^{(d)}-x_d\leq -a)\right]\\
&= \nu(x,(a,\infty)\times\mathbb{R}^{d-1})\mathbb{P}^x(X_0^{(d)}-x_d\leq -a)\\
&= 0,
\end{aligned}
$$

where we obtain lines 4 and 9 by set containment, line 5 by the fact that $A\cap B = A\backslash(A\cap B^c)$, line 10 by Lemma 3.4.6, and line 14 by Theorem 3.14. This contradiction gives us the desired result. □

Remark 3.4.4.

1. We can prove, by a similar technique, that PLOD implies condition (3.29). Thus, condition (3.29) in Feller processes yields a situations where PUOD and PLOD coincide.

2. Symbol $p(x,\xi)$ in the above theorem need not be bounded, only locally bounded.

Corollary 3.4.4. *For jump-Feller processes, i.e. $X\sim(b(x),0,\nu(x,dy))$ with bounded symbols $p(x,\xi)$, then condition (3.29), $\nu(x,(\mathbb{R}^d_+\cup\mathbb{R}^d_-)^c)=0$, is equivalent to X being associated, WA, PSA, PSD, POD, PUOD, PLOD in space.*

Proof. True by Theorems 3.15 and 3.17, and the implications between the different dependence forms given in Chapter 2. □

3.5 Applications and examples

Our above results are applicable to stochastically monotone Feller processes. We give a collection of interesting examples that satisfy stochastic monotonicity.

3.5.1 Lévy processes

Any Lévy process satisfies stochastic monotonicity. Let $(T_t)_{t\geq 0}$ be a semigroup of a Lévy process. Then, for $f \in \mathcal{F}_i$, we have

$$T_t f(x) = \mathbb{E}^x f(X_t) = \mathbb{E}^0 f(X_t + x).$$

Thus monotonicity of function f and of the expectation $\mathbb{E}^0$ gives us that $T_t f$ is a monotone function: $T_t f \in \mathcal{F}_i$.

Let $X = (X_t)_{t\geq 0}$ be a jump-Lévy process whose Lévy characteristics look like $(b, 0, \nu)$, where there is no state-space dependence. Then $\nu((\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0$ is equivalent to X_t being associated, PSA, PSD, and POD, since all Lévy processes are stochastically monotone. This was proven in Bauerle (2008) for association, PSD, and POD, but not for PSA, PUOD, and PLOD. Furthermore, the technique to prove condition (2.21) is equivalent to PSD and POD required Lévy copulas. Our method of short-time asymptotics avoids Lévy copulas altogether, and solely uses the Lévy measure.

Additionally, we have association, PSD, POD, PUOD, and PLOD in time and space since X has independent and stationary increments (see Theorem 3.6).

3.5.2 Ornstein-Uhlenbeck process

An Ornstein-Uhlenbeck process $X = (X_t)_{t\geq 0}$ in $\mathbb{R}^d$ is the solution to the general *Langevin equation*:

$$dX_t = -\lambda X_t dt + dL_t$$
$$X_0 = x \text{ a.s.}$$

where $\lambda > 0$, $L = (L_t)_{t\geq 0} \sim (b_L, \Sigma_L, \nu_L)$ is a Lévy process in $\mathbb{R}^d$, and $x \in \mathbb{R}^d$. Then OU-process looks like:

$$X_t = e^{-\lambda t}x + \int_0^t e^{-\lambda(t-s)}dL_t.$$

The semigroup $(T_t)_{t\geq 0}$ of this process is called the Mehler semigroup and is given by

$$T_t f(x) = \int_{\mathbb{R}^d} f(e^{\lambda t}x + y)\mu_t(dy), \quad L_t \sim \mu_t.$$

Claim 3.5.1. *The OU-process is stochastically monotone.*

Proof. Let $f \in B_b(\mathbb{R}^d)$ be an increasing function. Assume $x < y$, and fix some $t \geq 0$. Then $e^{-\lambda t}x < e^{-\lambda t}y$. This implies $f(e^{-\lambda t}x + z) \leq f(e^{-\lambda t}y + z)$ for all $z \in \mathbb{R}^d$. Hence,

$$T_t f(x) = \int_{\mathbb{R}^d} f(e^{-\lambda t}x + z)\mu_t(dz) \leq \int_{\mathbb{R}^d} f(e^{-\lambda t}y + z)\mu_t(dz) = T_t f(y).$$

Thus, $T_t f$ is an increasing function on $\mathbb{R}^d$. □

Process X has characteristic triplet: $(b_L - \lambda x, \Sigma_L, \nu_L)$ [3]. Thus, the characterization of positive dependence is equivalent to $\nu_L((\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0$.

3.5.3 Feller's pseudo-Poisson process

Here we construct a stochastically monotone pseudo-Poisson process. Let $S = (S(n))_{n\in\mathbb{N}}$ be a homogeneous Markov process taking values in $\mathbb{R}^d$. Let $(q^{(n)})_{n\in\mathbb{N}}$ define the n-step transition probabilities:

$$q^{(n)}(x, B) = \mathbb{P}(S(n) \in B | S(0) = x)$$

for all $B \in \mathcal{B}(\mathbb{R}^d)$. Let Q be the *transition operator* of S, defined by

$$(Qf)(x) = \int_{\mathbb{R}^d} f(y)q(x, dy)$$

for all $f \in B_b(\mathbb{R}^d)$, $x \in \mathbb{R}^d$. Note that $Q^n f(x) = \int_{\mathbb{R}^d} f(y) q^{(n)}(x, dy)$. Let $N = (N_t)_{t \geq 0}$ be a Poisson process with rate λ that is independent of S. Define $X = (X_t)_{t \geq 0}$ by subordination:

$$X_t := S(N_t) \text{ for all } t \geq 0.$$

Process X here Feller's pseudo-Poisson process, and is a Feller process. The semigroup $(T_t)_{t \geq 0}$ and generator $\mathcal{A}$ of X are given by:

$$T_t f(x) = e^{t[\lambda(Q-I)]} f(x) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} Q^n f(x),$$

$$\mathcal{A} f(x) = \lambda(Q - I) f(x) = \int_{\mathbb{R}^d} [f(y) - f(x)] \lambda q(x, dy).$$

Claim 3.5.2. *If S is a stochastically monotone Markov process, then X is stochastically monotone.*

Proof. We will show that for $f \in \mathcal{F}_i$, we have $T_t f \in \mathcal{F}_i$. Observe that by S stochastically monotone, we have that $q(x, B)$ is monotone function in x for all $B \in \mathcal{B}(\mathbb{R}^d)$ monotone set. Additionally, we have for $f \in B_b(\mathbb{R}^d) \cap \mathcal{F}_i$, $Qf(x) = \int_{\mathbb{R}^d} f(y) q(x, dy)$ is a monotone function. We want to show, by induction, that for all n, $G_n := e^{-\lambda t} \frac{(\lambda t)^n}{n!} Q^n f$ is a non-decreasing function.

Base Case: $n = 0$: $G_0(x) = e^{-\lambda t} f(x)$ is non-decreasing.

$n = 1$: $G_1(x) = e^{-\lambda t} \lambda t \, Qf(x) = e^{-\lambda t} \lambda t \int_{\mathbb{R}^d} f(z) q(x, dz)$ is non-decreasing.

Induction Hypothesis: Assume

$$G_n(x) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} Q^n f(x) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \int_{\mathbb{R}^d} f(z) q^{(n)}(x, dz)$$

is a non-decreasing function.

Inductive Step:

$$\begin{aligned} G_{n+1}(x) &= e^{-\lambda t} \frac{(\lambda t)^{n+1}}{(n+1)!} Q^{n+1} f(x) \\ &= e^{-\lambda t} \frac{(\lambda t)^{n+1}}{(n+1)!} \int_{\mathbb{R}^d} f(z) \, q^{(n+1)}(x, dz) \end{aligned}$$

$$= e^{-\lambda t}\frac{(\lambda t)^{n+1}}{(n+1)!}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} f(z)\, q^{(n)}(y,dz)\right) q(x,dy)$$
$$=: e^{-\lambda t}\frac{(\lambda t)^{n+1}}{(n+1)!}\int_{\mathbb{R}^d} H(y)q(x,dy)$$

where $H(y) = \int_{\mathbb{R}^d} f(z)q^{(n)}(y,dz)$ is a non-decreasing function in y by Induction Hypothesis, and line 3 is obtained by Chapman-Kolmogorov equation. Thus, by Base Case, the integral $\int_{\mathbb{R}^d} H(y)q(x,dy)$ is non-decreasing in x. Hence we get G_n is a non-decreasing function for all n. Hence, $T_t f$ is non-decreasing, giving us our desired result. □

Now to find the characteristic triplet $(b(x), \Sigma(x), \nu(x,dy))$, we consider the generator:

$$\begin{aligned}
\mathcal{A}f(x) &= \int_{\mathbb{R}^d}(f(z)-f(x))\lambda q(x,dz)\\
&= \int_{\mathbb{R}^d}(f(x+z)-f(x))\lambda q(x,dz+x)\\
&= \int_{\mathbb{R}^d}(f(x+z)-f(x))\lambda \hat{q}(x,dz), \quad \text{where } \hat{q}(x,B) := q(x,B+x)\\
&= \int_{\mathbb{R}^d}(f(x+z)-f(x)-\nabla f(x)\cdot z\mathbb{1}_{(0,1)}(|z|))\lambda\hat{q}(x,dz)\\
&\quad + \int_{\mathbb{R}^d}\nabla f(x)\cdot z\mathbb{1}_{(0,1)}(|z|))\lambda\hat{q}(x,dz)\\
&= \int_{\mathbb{R}^d}(f(x+z)-f(x)-\nabla f(x)\cdot z\mathbb{1}_{(0,1)}(|z|))\lambda\hat{q}(x,dz)\\
&\quad + \nabla f(x)\cdot\left(\int_{\mathbb{R}^d} z\mathbb{1}_{(0,1)}(|z|))\lambda\hat{q}(x,dz)\right).
\end{aligned}$$

Thus, the Lévy triplet will be $(b(x), \Sigma(x), \nu(x,dy))$, where

$$\begin{aligned}
b(x) &= \int_{\mathbb{R}^d} z\mathbb{1}_{(0,1)}(|z|))\lambda\hat{q}(x,dz)\\
\Sigma(x) &= 0\\
\nu(x,A) &= \lambda\hat{q}(x,A) = \lambda q(x,A+x).
\end{aligned}$$

3.5.4 Bochner's subordination of a Feller process

Consider a continuous-time Feller process $Y = (Y(t))_{t\geq 0}$ with semigroup $(T_t)_{t\geq 0}$ and generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$. Let $N = N_t$ be a subordinator independent of Y with Lévy characteristics (b, λ), i.e. has Lévy symbol

$$\eta(u) = ibu + \int_0^\infty (e^{iuy} - 1)\lambda(dy), \tag{3.41}$$

where $Ee^{iuN_t} = e^{t\eta(u)}$. Additionally, we can attain a Laplace transform of the subordinator, $\mathbb{E}e^{-uN_t} = e^{-t\psi(u)}$, where

$$\psi(u) := -\eta(iu) = bu + \int_0^\infty (1 - e^{-uy})\lambda(dy) \tag{3.42}$$

Function ψ is called the **Laplace symbol** or **Bernstein function** of the subordinator.

The following is a theorem of Phillips (see Applebaum [2]).

Theorem 3.18 (Phillips)**.** *Let $X = (X_t)_{t\geq 0}$ be given by the prescription $X_t = Y(N_t)$. Then X is a Feller process with semigroup $(T_t^X)_{t\geq 0}$ given by*

$$T_t^X f = \int_0^\infty (T_s f)\, \mu_{N_t}(ds).$$

and generator $(\mathcal{A}^X, \mathcal{D}(\mathcal{A}^X))$ given by

$$\mathcal{A}^X f = b\mathcal{A}f + \int_0^\infty (T_s f - f)\lambda(ds).$$

Claim 3.5.3. *If Y is a stochastically monotone Feller process with semigroup $(T_t)_{t\geq 0}$, i.e. $T_t f \in \mathcal{F}_i$ for $f \in \mathcal{F}_i$, and $N = (N_t)_{t\geq 0}$ is a subordinator, then $X = (X_t)_{t\geq 0}$ given by $X_t = Y(N_t)$ is a stochastically monotone Feller process.*

Proof. We already know that X is Feller with semigroup $(T_t^X)_{t\geq 0}$. So choose $f \in \mathcal{F}_i \cap C_b(\mathbb{R}^d)$. Then $T_s f \in \mathcal{F}_i \cap C_b(\mathbb{R}^d)$ for all $s \geq 0$. Choose $x < y$. Then $T_s f(x) \leq T_s f(y)$ for all $s \geq 0$.

Hence,

$$\begin{aligned} T_t^X f(x) &= \int_0^\infty (T_s f)(x)\, \mu_{N_t}(ds) \\ &\leq \int_0^\infty (T_s f)(y)\, \mu_{N_t}(ds) \\ &= T_t^X f(y). \end{aligned}$$

Thus, $T_t^X f \in \mathcal{F}_i$. □

Let Y now have characteristic triplet $(b(x), \Sigma(x), \nu(x, dy))$ and symbol $p(x, \xi)$. Then $X = Y(N)$ is a Feller process with symbol $p_X(x, \xi)$ that is given by

$$p_X(x, \xi) = \psi(p(x, \xi)) + \text{ lower order perturbation}$$

This *perturbation* is usually measured in a suitable scale of anisotropic function spaces. Bottcher, Schilling, and Wang use $H^n(\mathbb{R}^d) = W^{n,2}(\mathbb{R}^d)$ the classical L^2-Sobolev spaces [8].

Particularly interesting examples are when N is an α-stable subordinator, inverse Gaussian subordinator, and Gamma subordinator, and Y is a diffusion process $Y \sim (b(x), \Sigma(x), 0)$.

Example 3.5.1. Let Y be a stochastically monotone diffusion process in $\mathbb{R}^d$. This means Y has Lévy characteristics $(b(x), \Sigma(x), 0)$. Mu-fa Chen proved that such a process is stochastically monotone if and only if $\Sigma_{ij}(x)$ only depends on x_i and x_j, and $b_i(x) \leq b_i(y)$ whenever $x \leq y$ with $x_i = y_i$. The generator of Y would be given by:

$$\mathcal{A}^Y f(x) = b(x) \cdot \nabla f(x) + \frac{1}{2} \nabla \cdot \Sigma(x) \nabla f(x)$$

Let N be α-stable subordinator, thus having Lévy characteristics $(0, \lambda)$, where

$$\lambda(dy) = \frac{\alpha}{\Gamma(1-\alpha)} \frac{1}{y^{1+\alpha}} dy$$

The generator $\mathcal{A}^X$ of process $X = Y(N)$ looks like

$$\begin{aligned}\mathcal{A}^X f(x) &= \int_0^\infty (T_s f(x) - f(x))\lambda(ds) \\ &= \int_0^\infty (T_s f(x) - f(x))\frac{\alpha}{\Gamma(1-\alpha)}\frac{1}{s^{1+\alpha}}ds\end{aligned}$$

3.5.5 Lévy-driven stochastic differential equations

We investigate the stochastic monotonicity of Lévy-driven SDEs that are of the form

$$dX_t = c(X_{t-})dt + \sigma(X_{t-})dW_t + k \cdot dJ_t \tag{3.43}$$

where $W = (W_t)_{t\geq 0}$ is a standard Brownian motion, $J = (J_t)_{t\geq 0}$ is a Lévy process with characteristics $(0, 0, \nu_J)$, and $c : \mathbb{R}^d \to \mathbb{R}^{d\times n}$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times n}$ are bounded, Lipschitz continuous functions and $k \in \mathbb{R}^{d\times n}$ is a fixed matrix. Ideally, we wanted to look at SDEs of the form

$$dX_t = k(X_{t-})dJ_t.$$

where J is a Lévy (Poissonian) process. However, SDEs of this form often fail stochastic monotonicity even when coefficient k is a monotone function. For example, let J be an α-stable process. If k is non-constant, then X will not be stochastically monotone. For details of this example, see Wang [52].

We can still look at SDEs with constant coefficients in front of the Poissonian term. We examine their stochastic monotonicity below.

Let $X = (X_t)_{t\geq 0}$ be the solution process in $\mathbb{R}^d$ of the following SDE:

$$\begin{aligned}dX_t &= c(X_{t-})dt + \sigma(X_{t-})dW_t + k \cdot dJ_t \\ X_0 &= x \text{ a.s.}\end{aligned}$$

where $W = (W_t)_{t\geq 0} \sim (0, I, 0)$ is standard Brownian motion, $J = (J_t)_{t\geq 0} \sim (0, 0, \nu_J)$, $c : \mathbb{R}^d \to \mathbb{R}^{d\times n}$ is a monotone bounded, Lipschitz continuous function, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times n}$ is a monotone, bounded, Lipschitz continuous function such that $\sigma_{il}(x)$ only depends on x_i, for

all $l \in \{1, ..., n\}$, and $k \in \mathbb{R}^{d\times n}$. Then X is a rich Feller process with symbol

$$\begin{aligned}
-p(x,\xi) &= ic(x)\mathbf{1}\cdot\xi - \frac{1}{2}\xi\cdot\sigma(x)\sigma(x)^T\xi + \int_{\mathbb{R}^n}(e^{ik\xi\cdot z} - 1 - ik\xi\cdot z\mathbb{1}_{\hat{B}}(z))\nu_J(dz) \\
&= i\left(c(x)\mathbf{1} + \int_{|z|<1} kz(\mathbb{1}_{(0,1)}(|kz|) - \mathbb{1}_{(0,1)}(|z|))\nu_J(dz)\right)\cdot\xi \\
&\quad - \frac{1}{2}\xi\cdot\sigma(x)\sigma(x)^T\xi + \int_{\mathbb{R}^d}(e^{i\xi\cdot z} - 1 - i\xi\cdot z\mathbb{1}_{\hat{B}}(z))\nu(x,dz)
\end{aligned}$$

where $\nu(x,A) = \int_{\mathbb{R}^n}\mathbb{1}_A(kz)\nu_J(dz) = \nu_J(z\in\mathbb{R}^n : kz\in A)$, for all $A\in\mathcal{B}(\mathbb{R}^d\setminus\{0\})$, and $\mathbf{1} = (1, ..., 1)\in\mathbb{R}^n$

This yields the following characteristic triplet $(b(x), \Sigma(x), \nu(x, dy))$:

$$\begin{aligned}
b(x) &= c(x)\mathbf{1} + \int_{|z|<1} kz(\mathbb{1}_{(0,1)}(|kz|) - \mathbb{1}_{(0,1)}(|z|))\nu_J(dz) \\
\Sigma(x) &= \sigma(x)\sigma(x)^T \\
\nu(x,A) &= \nu_J(z\in\mathbb{R}^n : kz\in A) \quad A\in\mathcal{B}(\mathbb{R}^d\setminus\{0\})
\end{aligned}$$

Observe that the Lévy measure $\nu(x, dy)$ has no dependence on x and differs from Lévy measure ν_J by the transformation k. So we can write $N(dy) := \nu(x, dy)$ for all $x\in\mathbb{R}^d$. To show stochastic monotonicity, we employ the following comparison/monotonicity theorem given by Jie Ming Wang. [53].

Theorem 3.19 (Wang, [53]). *Let $X = (X_t)_{t\geq 0}$ be a rich Feller process in $\mathbb{R}^d$ with characteristics $(b(x), \Sigma(x), \nu(x, dy))$. Then X stochastically monotone if and only if the following three conditions hold:*

(i) For all i, j, $\Sigma_{ij}(x)$ only depends on variables x_i and x_j.

(ii) For all i and $n\geq 1$, whenever $x\leq y$ with $x_i = y_i$, we have

$$\begin{aligned}
&b_i(x) + \int_{\{|z_i|>1/n\}}(sgn(z_i)/n - z_i\mathbb{1}_{\hat{B}}(z)\nu(x,dz) + \int_{\{|z_i|\leq 1/n\}\cap\{|z|>1\}} z_i\nu(x,dz) \\
&\leq b_i(y) + \int_{\{|z_i|>1/n\}}(sgn(z_i)/n - z_i\mathbb{1}_{\hat{B}}(z)\nu(y,dz) + \int_{\{|z_i|\leq 1/n\}\cap\{|z|>1\}} z_i\nu(y,dz)
\end{aligned}$$

(iii) For all $x \leq y$ and A monotone increasing closed set,

$$\nu(x, A - x) \leq \nu(y, A - y) \quad \textit{if } y \notin A,$$
$$\nu(x, A^c - x) \geq \nu(y, A^c - y) \quad \textit{if } x \in A,$$

Note: The above result extends monotonicity of Gaussian component $\Sigma(x) = \{\Sigma_{ij}(x)\}_{1 \leq i,j \leq d}$ proven by Mu-Fa Chen [11].

Let's show our SDE with triplet $(b(x), \Sigma(x), \nu(x, dy))$ satisfies conditions (i), (ii), (iii).

Condition (i): For i, j, observe that $\Sigma_{ij}(x) = \sum_{l=1}^{n} \sigma_{il}(x)\sigma_{lj}^T(x)$. Then since $\sigma_{il}(x)$ only depends on x_i for all $l = 1, ..., n$, and $\sigma_{lj}^T(x) = \sigma_{jl}(x)$ only depends on x_j for all $l = 1, ..., n$, we have that $\Sigma_{ij}(x)$ only depends x_i, x_j.

Condition (ii): Choose $i \in \{1, ..., d\}$, $x \leq y$, with $x_i = y_i$, and $n \geq 1$. Observe that that the integral terms in Theorem (ii) will have no dependence on x for our process, i.e. $\nu(x, dz) = N(dz)$. Hence, the integral terms on both the left and right hand sides are equal. Now we just need to show $b_i(x) \leq b_i(y)$.

$$\begin{aligned} b_i(x) &= (a(x)\mathbf{1})_i + \int_{|z|<1} (kz)_i(\mathbb{1}_{(0,1)}(|kz|) - \mathbb{1}_{(0,1)}(|z|))\nu_J(dz) \\ &\leq (a(y)\mathbf{1})_i + \int_{|z|<1} (kz)_i(\mathbb{1}_{(0,1)}(|kz|) - \mathbb{1}_{(0,1)}(|z|))\nu_J(dz) \\ &= b_i(y) \end{aligned}$$

where $(b(x)\mathbf{1})_i = \sum_{j=1}^{n} a_{ij}(x)$, which is non-decreasing since a_{ij} are all non-decreasing functions.

Condition (iii): Choose A monotone increasing set and $x \leq y$. Then if $y \notin A$,

$$\begin{aligned} \nu(x, A - x) &= N(A - x) \\ &= \int_{\mathbb{R}^n} \mathbb{1}_{A-x}(kz)N(dz) \\ &= \int_{\mathbb{R}^n} \mathbb{1}_{A}(x + kz)N(dz) \\ &\leq \int_{\mathbb{R}^n} \mathbb{1}_{A}(y + kz)N(dz) \end{aligned}$$

$$
\begin{aligned}
&= N(A-y) \\
&= \nu(y, A-y)
\end{aligned}
$$

where we obtain the main inequality because $\mathbb{1}_A$ is a non-decreasing function. Hence $\mathbb{1}_A(x+kz) \leq \mathbb{1}_A(y+kz)$ for all $z \in \mathbb{R}^n$. Now if $x \in A$, then

$$
\begin{aligned}
\nu(x, A^c - x) &= N(A^c - x) \\
&= \int_{\mathbb{R}^n} \mathbb{1}_{A^c - x}(kz) N(dz) \\
&= \int_{\mathbb{R}^n} \mathbb{1}_{A^c}(x+kz) N(dz) \\
&\geq \int_{\mathbb{R}^n} \mathbb{1}_{A^c}(y+kz) N(dz) \\
&= N(A^c - y) \\
&= \nu(y, A^c - y)
\end{aligned}
$$

where we obtain the inequality because $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$ is a non-increasing function. Hence $\mathbb{1}_{A^c}(x+kz) \geq \mathbb{1}_{A^c}(y+kz)$ for all $z \in \mathbb{R}^n$. Thus, by Theorem 3.5.5, X is stochastically monotone.

So what would positive dependence, i.e. condition (3.29), look like for such a process? In other words, what does $\nu(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = \nu_J(y \in \mathbb{R}^n : ky \in (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0$ look like? Essentially, the entries of matrix k define linear boundaries of the concentration of Lévy measure ν_J. For example, if we take $n = d = 2$ and $k_{ij} > 0$ for all i, j (or $k_{ij} < 0$ for all i, j), then ν_J of Lévy process J is concentrated on $\{y : ky \in \mathbb{R}^2_+ \cup \mathbb{R}^2_-\}$. Observe that

$$
ky = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} k_{11}y_1 + k_{12}y_2 \\ k_{21}y_1 + k_{22}y_2 \end{bmatrix} =: \begin{bmatrix} (ky)_1 \\ (ky)_2 \end{bmatrix}
$$

Thus, we have two cases, either we have y such that both $(ky)_1, (ky)_2 \geq 0$ OR $(ky)_1, (ky)_2 \leq 0$. This implies the two sets of linear inequalities (A) and (B):

$$(A): \quad \begin{cases} k_{11}y_1 + k_{12}y_2 \geq 0 \\ k_{21}y_1 + k_{22}y_2 \geq 0. \end{cases}$$

and

$$(B): \quad \begin{cases} k_{11}y_1 + k_{12}y_2 \leq 0 \\ k_{21}y_1 + k_{22}y_2 \leq 0. \end{cases}$$

Thus, taking the union of our two cases (A) and (B), we can see the Lévy measure concentration in Figure 3.2, where the Lévy measure is 0 on the white space.

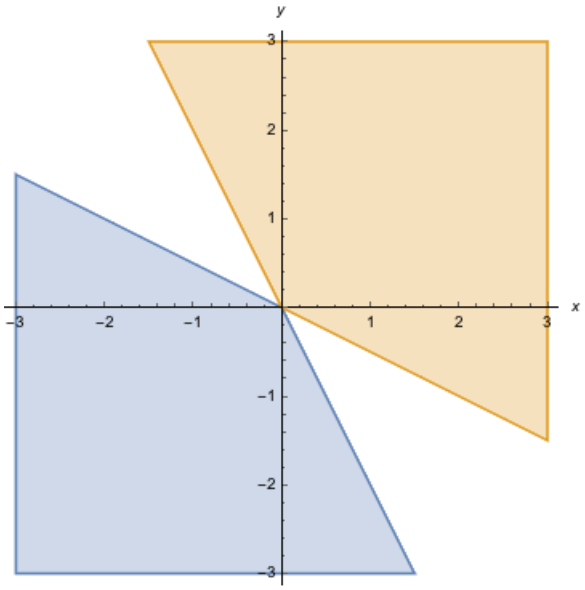

Figure 3.2: Transformed Lévy measure concentration

We chose entries of k to be all positive or all negative. But we don't need to make such a restriction, and different signs of k_{ij} will simply yield different linear inequalities, defining the new boundaries of concentration of the Lévy measure ν_J.

Remark 3.5.1. A very interesting application of the main results in this chapter actually arise in the case of time-inhomogeneous Markov processes. To study association in these time-inhomogeneous processes, one can transform the space into a larger space, in which

this newly transformed process is time-homogeneous in the new space. This is the subject of Chapter 4.

Chapter 4

Positive dependence in time-inhomogeneous Markov processes

We can also study dependence structures in time-inhomogeneous Markov processes. Time-inhomogeneous Markov processes arise as useful models in various applications, such as stochastic volatility models with jumps in financial modeling [13]. We will discuss what association means for a time-inhomogeneous Markov process and provide ways to characterize association using the time-dependent generators and the time-dependent symbols and characteristics.

4.1 Time-inhomogeneous Markov processes

Let $X = (X_t)_{t\geq 0}$ be a time-inhomogeneous Markov process on $\mathbb{R}^d$ on the space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$, i.e. X satisfies the Markov property:

$$\mathbb{P}(X_t \in A|\mathcal{G}_s) = \mathbb{P}(X_t \in A|X_s), \quad s \leq t, \quad A \in \mathcal{B}(\mathbb{R}^d) \tag{4.1}$$

and its **Markov evolution** $(T_{s,t})_{s\leq t}$ on $B_b(\mathbb{R}^d)$, defined by

$$T_{s,t}f(x) = \mathbb{E}(f(X_t)|X_s = x), \tag{4.2}$$

satisfies the properties of Proposition 3.1.1. Consider the Banach space $(C_0(\mathbb{R}^d), ||\cdot||_\infty)$. We say the Markov evolution is **strongly continuous** on $C_0(\mathbb{R}^d)$ if for every pair $0 \leq s \leq t < \infty$

$$\lim_{(u,v)\to(s,t)} ||T_{u,v}f - T_{s,t}f||_\infty = 0, \quad \forall f \in C_0(\mathbb{R}^d). \tag{4.3}$$

Markov evolutions that satisfy strong continuity (4.3) are called **Feller evolution systems (FES)**, and can be thought of as the time-inhomogeneous analogue to Feller semigroups studied in Chapter 3. For an FES, we can define a family of left and right generators. The right-generators $(\mathcal{A}_s^+)_{s\geq 0}$ of FES $(T_{s,t})_{s\leq t}$ is defined by

$$\mathcal{A}_s^+ f = \lim_{h\searrow 0} \frac{T_{s,s+h}f - f}{h} \tag{4.4}$$

for all $f \in \mathcal{D}(\mathcal{A}_s^+)$, the subspace of functions in $C_0(\mathbb{R}^d)$ for which the above limit exists in $||\cdot||_\infty$. Similarly, the left-generators $(\mathcal{A}_s^-, \mathcal{D}(\mathcal{A}_s^-))_{s\geq 0}$ by

$$\mathcal{A}_s^- f = \lim_{h\searrow 0} \frac{T_{s-h,s}f - f}{h}. \tag{4.5}$$

We can also express the left and right derivatives of the FES in terms of the left and right generators.

1. $\frac{d}{dt}^+ T_{s,t} = T_{s,t}\mathcal{A}_t^+$ (forward equation)
2. $\frac{d}{dt}^- T_{s,t} = T_{s,t}\mathcal{A}_t^-$
3. $\frac{d}{ds}^+ T_{s,t} = -\mathcal{A}_s^+ T_{s,t}$
4. $\frac{d}{ds}^- T_{s,t} = -\mathcal{A}_s^- T_{s,t}$ (backward equation).

Assume now that $C_c^\infty(\mathbb{R}^d) \subset \mathcal{D}(\mathcal{A}_s^+), \mathcal{D}(\mathcal{A}_s^-)$ for all $s \geq 0$. By the theorem of Courrège [14], we have that for every $s \geq 0$, $-\mathcal{A}_s^\pm|_{C_c^\infty(\mathbb{R}^d)}$ is a pseudo-differential operator:

$$\mathcal{A}_s^\pm|_{C_c^\infty(\mathbb{R}^d)} f(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{ix\cdot\xi} p_s^\pm(x,\xi)\hat{f}(\xi)d\xi \tag{4.6}$$

where $-p_s^{\pm}(x,\xi)$ is a cndf for each $s \geq 0$. We call $p_s^{\pm}(x,\xi)$ the **symbol of the generator** $\mathcal{A}_s^{\pm}$, and the $(p_s^{\pm}(x,\xi))_{s\geq 0}$ the **family of symbols of the process**. When $C_c^{\infty}(\mathbb{R}^d) \subset \mathcal{D}(\mathcal{A}_s^+), \mathcal{D}(\mathcal{A}_s^-)$ for all $s \geq 0$, we say that the generators have **rich domain** or that the associated **Markov process is rich**. In the FESs we study, the left and right generators will coincide. Böttcher gives conditions for this situation [7], which we write in the following theorem.

Theorem 4.1 (Böttcher, [7]). *Let $(T_{s,t})_{s\leq t}$ be a FES with left and right generators $(\mathcal{A}_s^+, \mathcal{D}(\mathcal{A}_s^+))_{s\geq 0}$ and $(\mathcal{A}_s^-, \mathcal{D}(\mathcal{A}_s^-))_{s\geq 0}$ with corresponding symbols $(p_s^+(x,\xi))_{s\geq 0}$ and $(p_s^-(x,\xi))_{s\geq 0}$. If*

$$p_s^{\pm}(x,\xi) \text{ is continuous in } s \text{ for all } x,\xi \in \mathbb{R}^d \tag{4.7}$$

and is bounded, i.e. there exists $C^{\pm} > 0$ such that

$$p_s^{\pm}(x,\xi) \leq C^{\pm}(1+|\xi|^2), \quad \forall s \geq 0, \ \forall x,\xi \in \mathbb{R}^d, \tag{4.8}$$

then $\mathcal{A}_s^+ = \mathcal{A}_s^-$ for all $s \geq 0$.

As a corollary to this theorem, conditions (4.7) and (4.8) give us just one family of generators and symbols to consider: $(\mathcal{A}_s)_{s\geq 0}$ and $(p_s(x,\xi))_{s\geq 0}$. Throughout this chapter, we often assume that $p_s(x,\xi)$ is **s-continuous** and **bounded**, i.e. satisfying (4.7) and (4.8), respectively.

Assume now a rich domain, i.e. $C_c^{\infty}(\mathbb{R}^d) \subset \mathcal{D}(\mathcal{A}_s)$ for all $s \geq 0$, and $p_s(x,\xi)$ is s-continuous and bounded. We have for each $s \geq 0$, an integro-differential operator $I(p_s)$ defined on $C_b^2(\mathbb{R}^d)$ by substituting the Lévy-Khintchine form

$$p_s(x,\xi) = ib_s(x)\cdot\xi - \frac{1}{2}\xi\cdot\Sigma_s(x)\xi + \int_{y\neq 0}(e^{i\xi\cdot y} - 1 - i\xi\cdot y\chi(y))\nu_s(x,dy) \tag{4.9}$$

into (4.6), and by elementary Fourier analysis,

$$I(p_s)f(x) = b_s(x)\cdot\nabla f(x) + \frac{1}{2}\nabla\cdot\Sigma_s(x)\nabla f(x) + \int_{y\neq 0}(f(x+y) - f(x) - \nabla f(x)\cdot y\chi(y))\nu_s(x,dy). \tag{4.10}$$

$I(p_s)$ clearly extends $\mathcal{A}_s$ onto $C_b^2(\mathbb{R}^d)$, and $I(p_s)|_{\mathcal{D}(\mathcal{A}_s)} = \mathcal{A}_s$.

4.1.1 Time-homogeneous transformation of time-inhomogeneous Markov process

Time-homogeneous Markov processes, such as the ones discussed in Chapter 3, have very nice properties and analytical tools. To take advantage of those tools in the time-inhomogeneous case, we can transform our time-inhomogeneous process X into a time-homogeneous process $\tilde{X}$ by adding another (deterministic) component to the process. We will outline the transformation of X to $\tilde{X}$ in this subsection. We follow the prescription used in Böttcher [7].

Let $X = (X_t)_{t\geq 0}$ be a time-inhomogeneous Markov process with sample space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$, state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, and Markov evolution $(T_{s,t})_{s\leq t}$, and corresponding Markov kernels $(P_{s,t})_{s\leq t}$ defined by

$$P_{s,t}(x, A) := T_{s,t}\mathbb{1}_A(x). \tag{4.11}$$

We define a transformed process in the following manner: To define the new sample space, let $\tilde{\Omega} := \mathbb{R}_+ \times \Omega$, where elements $\tilde{\omega} = (s, \omega)$, with $s \geq 0$, $\omega \in \Omega$. The σ-algebra will be $\tilde{\mathcal{G}}$, defined by

$$\tilde{\mathcal{G}} = \{A \subset \tilde{\Omega} : A_s \in \mathcal{G}, \ \forall s \geq 0\}, \tag{4.12}$$

where $A_s := \{\omega \in \Omega : (s, \omega) \in A\}$. The new state space will be defined to be $\mathbb{R}_+ \times \mathbb{R}^d$ with σ-algebra $\tilde{\mathcal{B}}$ defined by

$$\tilde{\mathcal{B}} = \{B \subset \mathbb{R}_+ \times \mathbb{R}^d : B_s \in \mathcal{B}(R^d), \ \forall s \geq 0\}, \tag{4.13}$$

where $B_s := \{x \in \mathbb{R}^d : (s, x) \in B\}$. From this, we can define a process $\tilde{X} = (\tilde{X}_t)_{t\geq 0}$ on $\mathbb{R}_+ \times \mathbb{R}^d$ by the prescription

$$\tilde{X}_t(\tilde{\omega}) = (s + t, X_{s+t}(\omega)) \tag{4.14}$$

where $\tilde{\omega} = (s, \omega)$. The family of probability measures $(\tilde{\mathbb{P}}^{\tilde{x}})_{\tilde{x}\in\mathbb{R}_+\times\mathbb{R}^d}$ is given by

$$\tilde{\mathbb{P}}^{\tilde{x}}(A|\tilde{X}_0 = \tilde{x}) = \tilde{\mathbb{P}}^{\tilde{x}}(A|\tilde{X}_0 = (s, x)) = \mathbb{P}(A_s|X_s = x), \quad A \in \tilde{\mathcal{G}} \tag{4.15}$$

From this we can define the transition kernel $(\tilde{P}_t)_{t\geq 0}$ by

$$\tilde{P}_t(\tilde{x}, B) := \tilde{\mathbb{P}}^{\tilde{x}}(\tilde{X}_t \in B | \tilde{X}_0 = \tilde{x}) = \mathbb{P}(X_{s+t} \in B_{s+t} | X_s = x), \quad B \in \tilde{\mathcal{B}} \tag{4.16}$$

Thus, this prescription has given us a process $\tilde{X} = (\tilde{X}_t)_{t\geq 0}$ with sample space $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{\mathbb{P}}^{\tilde{x}})_{\tilde{x}\in\mathbb{R}_+\times\mathbb{R}^d}$ and state space $(\mathbb{R}_+ \times \mathbb{R}^d, \tilde{\mathcal{B}})$.

Theorem 4.2 (Böttcher, [7])**.** *Stochastic process $\tilde{X}$ with sample space $(\tilde{\Omega}, \tilde{\mathcal{G}}, (\tilde{\mathcal{G}}_t)_{t\geq 0}, \tilde{\mathbb{P}}^{\tilde{x}})_{\tilde{x}\in\mathbb{R}_+\times\mathbb{R}^d}$ and state space $(\mathbb{R}_+ \times \mathbb{R}^d, \tilde{\mathcal{B}})$ given by the above prescription is a time-homogeneous Markov process, where $(\tilde{\mathcal{G}}_t)_{t\geq 0}$ is the natural filtration on $\tilde{X}$.*

Proof. See Böttcher [7]. □

(4.16) also gives us a manner to define the transition semigroup $(\tilde{T}_t)_{t\geq 0}$ on $(B_b(\mathbb{R}_+ \times \mathbb{R}^d), ||\cdot||_\infty)$ of Markov process $\tilde{X}$:

$$\tilde{T}_t f(\tilde{x}) = \tilde{\mathbb{E}}^{\tilde{x}} f(\tilde{X}_t) = \mathbb{E}(f_{s+t}(X_{s+t}) | X_s = x) = T_{s,s+t} f_{s+t}(x) \tag{4.17}$$

where $f_{s+t} : \mathbb{R}^d \to \mathbb{R}$ is defined by $f_{s+t}(x) := f(s+t, x)$. When given a time-inhomogeneous Markov process X on sample space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$ and state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we call the process $\tilde{X} = (\tilde{X}_t)_{t\geq 0}$ on sample space $(\tilde{\Omega}, \tilde{\mathcal{G}}, (\tilde{\mathcal{G}}_t)_{t\geq 0}, \tilde{\mathbb{P}}^{\tilde{x}})_{\tilde{x}\in\mathbb{R}_+\times\mathbb{R}^d}$, state space $(\mathbb{R}_+ \times \mathbb{R}^d, \tilde{\mathcal{B}})$, and semigroup $(\tilde{T}_t)_{t\geq 0}$ given by the above prescription the **transformed process of** X.

The following theorem of Böttcher shows that strong continuity of the FES $(T_{s,t})_{s\leq t}$ is preserved under the transformation.

Theorem 4.3 (Böttcher, [7])**.** *Let X be a time-inhomogeneous Markov process with Markov evolution $(T_{s,t})_{s\leq t}$, and let $\tilde{X}$ be the corresponding transformed process of X with semigroup $(\tilde{T}_t)_{t\geq 0}$. Then TFAE:*

- *$(T_{s,t})_{s\leq t}$ is a Feller evolution system on $C_0(\mathbb{R}^d)$.*
- *$(\tilde{T}_t)_{t\geq 0}$ is a Feller semigroup on $C_0(\mathbb{R}_+ \times \mathbb{R}^d)$.*

Proof. See Böttcher, [7]. □

Now that we have obtained a Feller process $\tilde{X}$ with Feller semigroup $(\tilde{T}_t)_{t\geq 0}$, we can ask what conditions on the process X will give us that $\tilde{X}$ is rich Feller? Under those conditions, what will the generator and extended generator of $\tilde{X}$ look like, and what will the symbol and characteristics look like?

Theorem 4.4 (Böttcher, [7]). *Let X be a time-inhomogeneous Markov process with FES $(T_{s,t})_{s\leq t}$, such that $C_c^\infty(\mathbb{R}^d) \subset \mathcal{D}(\mathcal{A}_s^+)$ for all $s \geq 0$, with symbols $p_s^+(x,\xi)$. Let $\tilde{X}$, the transformation of X, be the corresponding Feller process with semigroup $(\tilde{T}_t)_{t\geq 0}$, and $\tilde{\mathcal{A}}$ is the infinitesimal generator of $(\tilde{T}_t)_{t\geq 0}$, i.e. for all $f \in \mathcal{D}(\tilde{\mathcal{A}})$,*

$$\tilde{\mathcal{A}} f = \lim_{t \searrow 0} \frac{\tilde{T}_t f - f}{t}. \tag{4.18}$$

If $p_s^+(x,\xi)$ is s-continuous, then $\mathcal{D}(\tilde{\mathcal{A}}) \supset C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d)$.

We can also consider the extended pointwise generator L of process $\tilde{X}$.

Theorem 4.5 (Böttcher, [7]). *Let X be a time-inhomogeneous Markov process with FES $(T_{s,t})_{s\leq t}$, and let $\tilde{X}$ be the transformed process of X with Feller semigroup $(\tilde{T}_t)_{t\geq 0}$ and infinitesimal generator $\tilde{\mathcal{A}}$. Then the extended pointwise generator of $(\tilde{T}_t)_{t\geq 0}$ is defined for all $f \in C_0(\mathbb{R}_+ \times \mathbb{R}^d)$ such that*

(i) $f(\cdot, x) \in C^1(\mathbb{R}_+)$ for all $x \in \mathbb{R}^d$,

(ii) $f(s,\cdot) \in \mathcal{D}(\mathcal{A}_s^+)$ for all $s \geq 0$,

and is given by

$$Lf(\tilde{x}) := \frac{\partial}{\partial s} f(s,x) + \mathcal{A}_s^+ f_s(x), \tag{4.19}$$

where $\tilde{x} = (s,x)$. We also have $Lf(\tilde{x}) = \lim_{t\searrow 0} \frac{1}{t}(\tilde{T}_t f(\tilde{x}) - f(\tilde{x}))$ for all $\tilde{x}$.

Corollary 4.1.1. *Let $C_c^\infty(\mathbb{R}^d) \subset \mathcal{D}(\mathcal{A}_s^+)$ for all $s \geq 0$ with corresponding symbols $p_s^+(x,\xi)$. Then L, given by (4.19), is defined for all $f \in C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d)$ and is a pseudo-differential operator with symbol*

$$p^L(\tilde{x},\tilde{\xi}) = ir + p_s^+(x,\xi), \quad \tilde{x} = (s,x), \ \tilde{\xi} = (r,\xi) \tag{4.20}$$

Remark 4.1.1. Again, if we assume the $p_s^{\pm}(x,\xi)$ is s-continuous and bounded in the sense of (4.7) and (4.8), then we could just consider the generators and symbols of X without the $+$ superscript.

Corollary 4.1.2. *Assume $C_c^\infty(\mathbb{R}^d) \subset \mathcal{D}(\mathcal{A}_s)$ for all $s \geq 0$ and the symbols $p_s(x,\xi)$ are s-continuous and bounded. Then process $\tilde{X}$ has extended pointwise generator*

$$Lf(\tilde{x}) = \frac{\partial}{\partial s} f(s,x) + \mathcal{A}_s f_s(x) = \frac{\partial}{\partial s} f(s,x) + I(p_s) f_s(x) \tag{4.21}$$

where $I(p_s)$ defined by (4.10), and $L|_{\mathcal{D}(\tilde{\mathcal{A}})} = \tilde{\mathcal{A}}$.

Now recall that a Feller process has an extended generator that is an integro-differential operator in the sense (3.23) and has a corresponding symbol and characteristics. We would like to compute these for the Feller process $\tilde{X}$ above. We can do this using the RHS of extended pointwise generator L. Let $I(\tilde{p})$ be the integro-differential operator defined on $C_b^2(\mathbb{R}_+ \times \mathbb{R}^d)$. Then $I(\tilde{p})$ extends $\tilde{\mathcal{A}}$, and $I(\tilde{p})|_{\mathcal{D}(\tilde{\mathcal{A}})} = \tilde{\mathcal{A}} = L|_{\mathcal{D}(\tilde{\mathcal{A}})}$. Then using the RHS of (4.21) and (4.10), we have

$$\begin{aligned}
I(\tilde{p})f(\tilde{x}) &= \frac{\partial}{\partial s} f(s,x) + I(p_s) f_s(x) \\
&= \frac{\partial}{\partial s} f(s,x) + b_s(x) \cdot \nabla f_s(x) + \frac{1}{2} \nabla \cdot \Sigma_s(x) \nabla f_s(x) \\
&\quad + \int_{y \neq 0} (f_s(x+y) - f_s(x) - \nabla f_s(x) \cdot y\chi(y)) \nu_s(x, dy) \\
&= \tilde{b}(s,x) \cdot \nabla f(s,x) + \frac{1}{2} \nabla \cdot \tilde{\Sigma}(s,x) \nabla f(s,x) \\
&\quad + \int_{y \neq 0} (f(s,x+y) - f(s,x) - \nabla f(s,x) \cdot y\chi(y)) \nu_s(x, dy) \\
&= \tilde{b}(\tilde{x}) \cdot \nabla f(\tilde{x}) + \frac{1}{2} \nabla \cdot \tilde{\Sigma}(\tilde{x}) \nabla f(\tilde{x}) \\
&\quad + \int_{\tilde{y} \neq 0} (f(\tilde{x}+\tilde{y}) - f(\tilde{x}) - \nabla f(\tilde{x}) \cdot \tilde{y}\chi(\tilde{y})) \tilde{\nu}(\tilde{x}, d\tilde{y})
\end{aligned}$$

where $\tilde{b} : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ defined by

$$\tilde{b}(\tilde{x}) = \tilde{b}(s,x) = (1, b_s(x)),$$

$\tilde{\Sigma} : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1\times d+1}$ defined by

$$\begin{aligned}
\tilde{\Sigma}^{i0}(\tilde{x}) &= 0, \ \ \forall j = 0, ..., d \\
\tilde{\Sigma}^{0j}(\tilde{x}) &= 0, \ \ \forall i = 0, ..., d \\
\tilde{\Sigma}^{ij}(\tilde{x}) &= \Sigma_s^{ij}(x) \ \ \forall i, j = 1, ..., d
\end{aligned}$$

and $\tilde{\nu}(\tilde{x}, d\tilde{y})$ is a Lévy measure on $\mathcal{B}(\mathbb{R}^{d+1} \setminus \{0\})$ given by

$$\tilde{\nu}(\tilde{x}, d\tilde{y}) = \nu_s(x, dy)\delta_0(dr),$$

where $\tilde{y} = (r, y)$, and δ_0 is Dirac measure at 0.

We should note that $\tilde{\Sigma}(\tilde{x})$ is a symmetric, positive semidefinite matrix in $d + 1 \times d + 1$, by the symmetric, positive semidefiniteness of $\Sigma_s(x)$. Also, $\tilde{\nu}(\tilde{x}, d\tilde{y})$ is a Lévy measure on $\mathcal{B}(\mathbb{R}^{d+1} \setminus \{0\})$:

$$\begin{aligned}
\int_{\tilde{y}\neq 0} (1 \wedge |\tilde{y}|^2)\tilde{\nu}(\tilde{x}, d\tilde{y}) &= \int_{[0,\infty)} \int_{y\neq 0} (1 \wedge |(r, y)|^2)\nu_s(x, dy)\delta_0(dr) \\
&= \int_{y\neq 0} (1 \wedge |(0, y)|^2)\nu_s(x, dy) \\
&< \infty
\end{aligned}$$

This triplet $(\tilde{b}(\tilde{x}), \tilde{\Sigma}(\tilde{x}), \tilde{\nu}(\tilde{x}, d\tilde{y}))$ forms the characteristic triplet of $\tilde{X}$. Hence, we have the following corollary.

Corollary 4.1.3. *Given X and transformed process $\tilde{X}$, where $C_c^\infty(\mathbb{R}^d) \subset \mathcal{D}(\mathcal{A}_s)$ for all $s \geq 0$ and $p_s(x, \xi)$ is s-continuous and bounded, then $\tilde{X}$ has characteristic triplet $(\tilde{b}(\tilde{x}), \tilde{\Sigma}(\tilde{x}), \tilde{\nu}(\tilde{x}, d\tilde{y}))$ given above, and integro-differential operator $I(\tilde{p})$, defined by equations*

$$I(\tilde{p})f(\tilde{x}) = \tilde{b}(\tilde{x}) \cdot \nabla f(\tilde{x}) + \frac{1}{2}\nabla \cdot \tilde{\Sigma}(\tilde{x})\nabla f(\tilde{x}) + \int_{\tilde{y}\neq 0} (f(\tilde{x} + \tilde{y}) - f(\tilde{x}) - \nabla f(\tilde{x}) \cdot \tilde{y}\chi(\tilde{y}))\tilde{\nu}(\tilde{x}, d\tilde{y}) \tag{4.22}$$

or

$$I(\tilde{p})f(\tilde{x}) = \frac{\partial}{\partial s} f(s, x) + I(p_s)f_s(x) \tag{4.23}$$

is an extension of generator L and $\tilde{\mathcal{A}}$ onto $C_b^2(\mathbb{R}^d)$, i.e.

$$I(\tilde{p})|_{\mathcal{D}(L)} = L, \qquad I(\tilde{p})|_{\mathcal{D}(\tilde{\mathcal{A}})} = \tilde{\mathcal{A}}. \tag{4.24}$$

Also, the symbol of process $\tilde{X}$ has representations

$$\tilde{p}(\tilde{x},\tilde{\xi}) = i\tilde{b}(\tilde{x})\cdot\tilde{\xi} - \frac{1}{2}\tilde{\xi}\cdot\tilde{\Sigma}(\tilde{x})\tilde{\xi} + \int_{\tilde{y}\neq 0}(e^{i\tilde{\xi}\cdot\tilde{y}} - 1 - i\tilde{\xi}\cdot\tilde{y}\chi(\tilde{y}))\tilde{\nu}(\tilde{x},d\tilde{y}) \tag{4.25}$$

or

$$\tilde{p}(\tilde{x},\tilde{\xi}) = ir + p_s(x,\xi), \qquad \tilde{x} = (s,x),\ \ \tilde{\xi} = (r,\xi). \tag{4.26}$$

Finally, the boundedness and s-continuity of $p_s(x,\xi)$ of process X yields boundedness of $\tilde{p}(\tilde{x},\tilde{\xi})$ as was given in Chapter 3.

Theorem 4.6 (Böttcher, [7])**.** *Let X be the time-inhomogeneous Markov process with FES $(T_{s,t})_{s\leq t}$, generators $(\mathcal{A}_s)_{s\geq 0}$ with rich domains, and $p_s(x,\xi)$ is s-continuous and bounded as in (4.7) and (4.8). If $C_c^\infty(\mathbb{R}^d)$ is the core of $\mathcal{A}_s$ for all $s \geq 0$, then Feller process $\tilde{X}$ has symbol $\tilde{p}(\tilde{x},\tilde{\xi})$ that is bounded, i.e. there exists $C > 0$ such that*

$$\sup_{\tilde{x}\in\mathbb{R}_+\times\mathbb{R}^d} |\tilde{p}(\tilde{x},\tilde{\xi})| \leq C(1+|\tilde{\xi}|^2), \tag{4.27}$$

for all $\tilde{\xi}\in\mathbb{R}_+\times\mathbb{R}^d$.

4.2 Association of time-inhomogeneous Markov processes

Now that we have discussed analytical tools of strongly continuous, time-inhomogeneous Markov processes, we can characterize the association of such processes. We would first like to define what spatial association means in terms of the Feller evolution system. Our definition extends from Lindqvist's definition of association of time-inhomogeneous Markov chains in discrete time and on a finite partially ordered space [33].

Definition 4.1. Let $X = (X_t)_{t\geq 0}$ be a time-inhomogeneous Markov process with Markov evolution $(T_{s,t})_{s\leq t}$. We say X is **spatially associated** if for all $s \leq t$, $f, g \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i$,

$$T_{s,t}fg \geq T_{s,t}f \, T_{s,t}g \tag{4.28}$$

Definition 4.2. Let $X = (X_t)_{t\geq 0}$ be a time-inhomogeneous Markov process with Markov evolution $(T_{s,t})_{s\leq t}$. We say X is **temporally associated** if for all $0 \leq t_1 < ... < t_n$, $(X_{t_1}, ..., X_{t_n})$ is associated in $\mathbb{R}^{dn}$.

Definition 4.3. Let $X = (X_t)_{t\geq 0}$ be a time-inhomogeneous Markov process with Markov evolution $(T_{s,t})_{s\leq t}$. We say X is **stochastically monotone** if for all $s \leq t$, $f \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i$,

$$T_{s,t}f \in \mathcal{F}_i. \tag{4.29}$$

Note that the definition of spatial association is stronger than the statement "X_t is associated for all $t \geq 0$." Essentially, (4.28) means X_t, conditioned on $X_s = x$, is associated, for all $x \in \mathbb{R}^d$, for all $s \leq t$. Such a definition is more useful in applications. For example, see [33] for an application in reliability theory.

We give a characterization of spatial association for strongly continuous, time-inhomogeneous Markov processes based on the generators $\mathcal{A}_s$. We apply this to characterize association of such processes of the jump variety, i.e. $(b_s(x), 0, \nu_s(x, dy))$.

Theorem 4.7 (Tu, 2017b, [51])**.** *Let $X = (X_t)_{t\geq 0}$ be a strongly continuous, time-inhomogeneous Markov process with Feller evolution system $(T_{s,t})_{s\leq t}$, generators $(\mathcal{A}_s)_{s\geq 0}$ with rich domains, and that $C_c^\infty(\mathbb{R}^d)$ is the core for $\mathcal{A}_s$, for all $s \geq 0$. Let the corresponding symbols $p_s(x, \xi)$ be s-continuous and bounded as in (4.7) and (4.8), and $I(p_s)$ be the integro-differential operator that is the extended generator onto $C_b^2(\mathbb{R}^d)$. If X is stochastically monotone, then X is spatially associated if and only if*

$$I(p_s)fg \geq fI(p_s)g + gI(p_s)f, \quad \forall f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i \tag{4.30}$$

Proof. Let $\tilde{X} = (\tilde{X}_t)_{t\geq 0}$ on $\mathbb{R}_+ \times \mathbb{R}^d$ be transformation of X, as prescribed in Section 4.1.1, which has Feller semigroup $(\tilde{T}_t)_{t\geq 0}$, generator $(\tilde{\mathcal{A}}, \mathcal{D}(\tilde{\mathcal{A}}))$ with rich domain, bounded symbol

$\tilde{p}(\tilde{x},\tilde{\xi})$, characteristics $(\tilde{b}(\tilde{x}),\tilde{\Sigma}(\tilde{x}),\tilde{\nu}(\tilde{x},d\tilde{y}))$ and extended generator $I(\tilde{p})$ on $C_b^2(\mathbb{R}_+\times\mathbb{R}^d)$, as given to us by Theorem 4.6 and Corollary 4.1.3.

($\Rightarrow$). Assume $T_{s,s+t}fg \geq T_{s,s+t}fT_{s,s+t}g$ for all $s,t\geq 0$ and all $f,g\in C_b(\mathbb{R}^d)\cap\mathcal{F}_i$. Choose $h,k\in C_b(\mathbb{R}_+\times\mathbb{R}^d)\cap\mathcal{F}_i$, $\tilde{x}=(s,x)$. Then $h_{s+t},k_{s+t}\in C_b(\mathbb{R}^d)\cap\mathcal{F}_i$, and

$$\begin{aligned}\tilde{T}_thk(\tilde{x}) &= T_{s,s+t}h_{s+t}k_{s+t}(x)\\ &\geq T_{s,s+t}h_{s+t}(x)\cdot T_{s,s+t}k_{s+t}(x)\\ &= \tilde{T}_th(\tilde{x})\cdot\tilde{T}_tk(\tilde{x}).\end{aligned}$$

Observe that the bounded symbol $\tilde{p}(\tilde{x},\tilde{\xi})$ also satisfies $\tilde{x}\mapsto\tilde{p}(\tilde{x},0)$ is continuous, since

$$\tilde{p}(\tilde{x},0) = i\cdot(0)+p_s(x,0)=0.$$

So by Theorem 3.16, we have that the extended generator $I(\tilde{p})$ satisfies

$$I(\tilde{p})hk \geq hI(\tilde{p})k + kI(\tilde{p})h, \qquad h,k\in C_b^2(\mathbb{R}_+\times\mathbb{R}^d)\cap\mathcal{F}_i. \tag{4.31}$$

Choose $f,g\in C_b(\mathbb{R}^d)\cap\mathcal{F}_i$. Then there exists $h,k\in C_b(\mathbb{R}_+\times\mathbb{R}^d)\cap\mathcal{F}_i$, where h,k are constant with respect to the first argument, and $f(x)=h(\tilde{x})$ and $g(x)=k(\tilde{x})$. Choose $\tilde{x}=(s,x)$. Then

$$\begin{aligned}I(\tilde{p})hk(\tilde{x}) &= \frac{\partial}{\partial s}h(s,x)k(s,x)+I(p_s)h_sk_s(x)\\ &= 0+I(p_s)fg(x)\\ &= I(p_s)fg(x)\end{aligned}$$

and

$$\begin{aligned}&h(\tilde{x})I(\tilde{p})k(\tilde{x})+k(\tilde{x})I(\tilde{p})h(\tilde{x})\\ &= h(s,x)\left(\frac{\partial}{\partial s}k(s,x)+I(p_s)k_s(x)\right)+k(s,x)\left(\frac{\partial}{\partial s}h(s,x)+I(p_s)h_s(x)\right)\end{aligned}$$

$$= h(\tilde{x})I(p_s)k_s(x) + k(\tilde{x})I(p_s)h_s(x)$$
$$= f(x)I(p_s)g(x) + g(x)I(p_s)f(x).$$

Thus, by (4.31), we have

$$I(p_s)fg \geq fI(p_s)g + gI(p_s)f.$$

($\Leftarrow$). Assume, for all $s \geq 0$, $I(p_s)fg \geq fI(p_s)g + gI(p_s)f$, $\forall f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$. Choose $f, g \in C_b^2(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{F}_i$, $\tilde{x} = (s, x)$, then

$$\begin{aligned} I(\tilde{p})fg(\tilde{x}) &= \frac{\partial}{\partial s} f(s,x)g(s,x) + I(p_s)f_s g_s(x) \\ &= f(s,x)\frac{\partial}{\partial s} g(s,x) + g(s,x)\frac{\partial}{\partial s} f(s,x) + I(p_s)f_s g_s(x) \\ &\geq f(s,x)\frac{\partial}{\partial s} g(s,x) + g(s,x)\frac{\partial}{\partial s} f(s,x) + f_s(x)I(p_s)g_s(x) + g_s(x)I(p_s)f_s(x) \quad (4.32) \\ &= f(s,x)\left(\frac{\partial}{\partial s} g(s,x) + I(p_s)g_s(x)\right) + g(s,x)\left(\frac{\partial}{\partial s} f(s,x) + I(p_s)f_s(x)\right) \\ &= f(\tilde{x})I(\tilde{p})g(\tilde{x}) + g(\tilde{x})I(\tilde{p})f(\tilde{x}). \end{aligned}$$

Note that we assumed $(T_{s,t})_{s \leq t}$ is stochastically monotone. However, this does not imply that $(\tilde{T}_t)_{t \geq 0}$ is stochastcally monotone. To see this, choose $\tilde{x} = (s, x)$ and $\tilde{y} = (r, y)$, where $\tilde{x} \leq \tilde{y}$ with $s < r$. Then let $f \in \mathcal{F}_i \cap C_b(\mathbb{R}_+ \times \mathbb{R}^d)$. Observe that

$$\begin{aligned} \tilde{T}_t f(\tilde{x}) &= \tilde{\mathbb{E}}^{\tilde{x}} f(\tilde{X}_t) \\ &= \mathbb{E}(f_{s+t}(X_{s+t}) | X_s = x) \\ &\not\leq \mathbb{E}(f_{r+t}(X_{r+t}) | X_r = y) \\ &= \tilde{T}_t f(\tilde{y}) \end{aligned}$$

since the sample paths of X may not be monotone non-decreasing. But we can still get our desired result from the stochastic monotonicity of $(T_{s,t})_{s \leq t}$. Choose $h, k \in C_b^2(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{F}_i$. Then $\tilde{T}_t h|_{\{s\} \times \mathbb{R}^d}$, $\tilde{T}_t k|_{\{s\} \times \mathbb{R}^d}$, $\tilde{T}_t hk|_{\{s\} \times \mathbb{R}^d} \in C_b^2(\{s\} \times \mathbb{R}^d) \cap \mathcal{F}_i$ for a fixed $s \geq 0$. It is easy to see that these functions will be in $C_b^2(\{s\} \times \mathbb{R}^d)$. To see that they are non-decreasing on

$\{s\} \times \mathbb{R}^d$, choose $\tilde{x} := (s, x) \leq (s, y) =: \tilde{y}$. Then

$$\begin{aligned}\tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) = \tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}(s, x) &= T_{s,s+t} h_{s+t}(x) \\ &\leq T_{s,s+t} h_{s+t}(y) \\ &= \tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}(\tilde{y})\end{aligned}$$

by stochastic monotonicity of $(T_{r,t})_{r\leq t}$. Observe that there exists $v \in C_b^2(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{F}_i$ such that v is constant with respect to the first argument in $\mathbb{R}_+$ and $v(s, x) = \tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}(s, x)$. Similarly, there is $w \in C_b^2(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{F}_i$ such that w is constant with respect to first argument, and $w(s, x) = \tilde{T}_t k|_{\{s\}\times\mathbb{R}^d}(s, x)$. By inequality (4.32), we have

$$I(\tilde{p})vw \geq vI(\tilde{p})w + wI(\tilde{p})v,$$

which implies for any $x \in \mathbb{R}^d$, with $\tilde{x} = (s, x)$,

$$\begin{aligned}I(\tilde{p})\left(\tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}\tilde{T}_t k|_{\{s\}\times\mathbb{R}^d}\right)(\tilde{x}) \geq{}& \tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) \cdot I(\tilde{p})\tilde{T}_t k|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) \\ &+ \tilde{T}_t k|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) \cdot I(\tilde{p})\tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}(\tilde{x})\end{aligned} \tag{4.33}$$

Now define $F : [0, \infty) \to C_b(\mathbb{R}_+ \times \mathbb{R}^d)$, by

$$F(t) := \tilde{T}_t hk - \tilde{T}_t h \cdot \tilde{T}_t k. \tag{4.34}$$

Define $G : [0, \infty) \to C_b(\mathbb{R}_+ \times \mathbb{R}^d)$, by

$$G(t) := F'(t) - I(\tilde{p})F(t). \tag{4.35}$$

It is not hard to verify that F, G are continuous on $[0, \infty)$ with respect to local uniform convergence. By Theorem 3.4.5, we have the solution

$$F(t) = \tilde{T}_t F(0) + \int_0^t \tilde{T}_{t-r} G(r)dr = \int_0^t \tilde{T}_{t-r} G(r)dr. \tag{4.36}$$

Now, choose $\tilde{x} = (s, x)$. Then by (4.33)

$$\begin{aligned}
F'(t)(\tilde{x}) &= I(\tilde{p})\tilde{T}_t hk(\tilde{x}) - (\tilde{T}_t h(\tilde{x}) \cdot I(\tilde{p})\tilde{T}_t k(\tilde{x}) + \tilde{T}_t k(\tilde{x}) \cdot I(\tilde{p})\tilde{T}_t h(\tilde{x})) \\
&= I(\tilde{p})\tilde{T}_t hk|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) - \Big(\tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) \cdot I(\tilde{p})\tilde{T}_t k|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) \\
&\qquad\qquad + \tilde{T}_t k|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) \cdot I(\tilde{p})\tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}(\tilde{x})\Big) \\
&\geq I(\tilde{p})\tilde{T}_t hk|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) - I(\tilde{p})\left(\tilde{T}_t h|_{\{s\}\times\mathbb{R}^d}\tilde{T}_t k|_{\{s\}\times\mathbb{R}^d}\right)(\tilde{x}) \\
&= I(\tilde{p})F(t)|_{\{s\}\times\mathbb{R}^d}(\tilde{x}) \\
&= I(\tilde{p})F(t)(\tilde{x}).
\end{aligned}$$

Thus, $G(t)(\tilde{x}) = F'(t)(\tilde{x}) - I(\tilde{p})F(t)(\tilde{x}) \geq 0$. In other words, $G(t)|_{\{s\}\times\mathbb{R}^d} \geq 0$. Hence,

$$F(t)|_{\{s\}\times\mathbb{R}^d} = \int_0^t \tilde{T}_{t-r} G(r)|_{\{s\}\times\mathbb{R}^d} dr \geq 0. \tag{4.37}$$

This finally yields $\tilde{T}_t hk(\tilde{x}) \geq \tilde{T}_t h(\tilde{x}) \cdot \tilde{T}_t k(\tilde{x})$, for all $\tilde{x} \in \{s\} \times \mathbb{R}^d$, which then yields

$$T_{s,s+t} h_{s+t} k_{s+t}(x) \geq T_{s,s+t} h_{s+t}(x) \cdot T_{s,s+t} k_{s+t}(x) \tag{4.38}$$

for all $x \in \mathbb{R}^d$. Now let $f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$. Then there are functions $h, k \in C_b^2(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{F}_i$ that are constant with respect to the first argument, such that $f(x) = h(\tilde{x})$ and $g(x) = k(\tilde{x})$. Then by (4.38), we have

$$T_{s,s+t} fg(x) \geq T_{s,s+t} f(x) \cdot T_{s,s+t} g(x). \tag{4.39}$$

Note that we chose a fixed arbitrary $s \geq 0$. We could follow the above procedure using any $s \geq 0$, and thus we would obtain (4.39) for all $s, t \geq 0$, giving us our desired result. □

We can now apply this to characterize strongly continuous, time-inhomogeneous Markov jump-processes based on the time-dependent Lévy measures.

Theorem 4.8 (Tu, 2017b, [51]). *Let $X = (X_t)_{t\geq 0}$ be a strongly continuous, time-inhomogeneous Markov process with Feller evolution system $(T_{s,t})_{s\leq t}$, generators $(\mathcal{A}_s)_{s\geq 0}$ with rich domains, and that $C_c^\infty(\mathbb{R}^d)$ is the core for $\mathcal{A}_s$, for all $s \geq 0$. Let the corresponding*

symbols $p_s(x,\xi)$ be s-continuous and bounded as in (4.7) and (4.8) with characteristic triplet $(b_s(x), 0, \nu_s(x,dy))$. If X is stochastically monotone, then X is spatially associated if and only if

$$\nu_s(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0, \quad \forall s \geq 0,\ x \in \mathbb{R}^d. \tag{4.40}$$

Proof. ($\Leftarrow$) Sufficient condition. Assume (4.40). Let $I(p_s)$ be the extended generator onto $C_b^2(\mathbb{R}^d)$, which is an integro-differential operator. Choose $s \geq 0$, $f, g \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$. Then

$$\begin{aligned}
&I(p_s)fg(x) - f(x)I(p_s)g(x) - g(x)I(p_s)f(x) \\
&= \int_{\mathbb{R}^d \setminus \{0\}} (f(x+y) - f(x))(g(x+y) - g(x))\nu_s(x,dy) \\
&= \int_{\mathbb{R}^d_+ \setminus \{0\}} (f(x+y) - f(x))(g(x+y) - g(x))\nu_s(x,dy) \\
&\quad + \int_{\mathbb{R}^d_- \setminus \{0\}} (f(x+y) - f(x))(g(x+y) - g(x))\nu_s(x,dy) \\
&\geq 0.
\end{aligned}$$

Then by Theorem 4.7, X is spatially associated.

($\Rightarrow$) Necessary condition. We just show the proof for dimension $d = 2$. Let X be spatially associated. Then by Theorem 4.7, $I(p_s)fg \geq fI(p_s)g + gI(p_s)f$ for all $s \geq 0$, $f, g \in C_b^2(\mathbb{R}^2) \cap \mathcal{F}_i$. This implies

$$\int_{\mathbb{R}^2 \setminus \{0\}} (f(x+y) - f(x))(g(x+y) - g(x))\nu_s(x,dy) \geq 0, \quad \forall s \geq 0.$$

Assume for contradiction that there exists $t_0 \geq 0$ and $x = (x_1, x_2) \in \mathbb{R}^2$ such that $\nu_{t_0}(x, (\mathbb{R}^2_+ \cup \mathbb{R}^2_-)^c) > 0$. WLOG, say $\nu_{t_0}(x, (0,\infty) \times (-\infty, 0)) > 0$. Then by continuity of measure, there exists $a > 0$ such that $\nu_{t_0}(x, (a,\infty) \times (-\infty, -a)) > 0$. Fix $\epsilon > 0$ and choose $f, g \in C_b^\infty(\mathbb{R}^2) \cap \mathcal{F}_i$ such that

$$f(y_1, y_2) = \begin{cases} 0 & \text{if} \quad y_1 \leq x_1 + \epsilon a \\ 1 & \text{if} \quad y_1 \geq x_1 + a, \end{cases}$$

$$g(y_1, y_2) = \begin{cases} 0 & \text{if} \quad y_2 \geq x_2 - \epsilon a \\ -1 & \text{if} \quad y_2 \leq x_2 - a. \end{cases}$$

This implies $f(x) = g(x) = 0$. Hence,

$$\begin{aligned}
0 &\leq \int_{y \neq 0} (f(x+y) - f(x))(g(x+y) - g(x))\nu_{t_0}(x, dy) \\
&= \int_{y \neq 0} f(x+y)g(x+y)\nu_{t_0}(x, dy) \\
&= \int_{(a,\infty)\times(-\infty,-a)} f(x+y)g(x+y)\nu_{t_0}(x, dy) + \int_{(a,\infty)\times[-a,-\epsilon a]} f(x+y)g(x+y)\nu_{t_0}(x, dy) \\
&\quad + \int_{[\epsilon a,a]\times(-\infty,-a)} f(x+y)g(x+y)\nu_{t_0}(x, dy) + \int_{[\epsilon a,a]\times[-a,-\epsilon a]} f(x+y)g(x+y)\nu_{t_0}(x, dy) \\
&= -\nu_{t_0}(x, (a,\infty) \times (-\infty, -a)) - \int_{(a,\infty)\times[-a,-\epsilon a]} g(x+y)\nu_{t_0}(x, dy) \\
&\quad + \int_{[\epsilon a,a]\times(-\infty,-a)} f(x+y)\nu_{t_0}(x, dy) + \int_{[\epsilon a,a]\times[-a,-\epsilon a]} f(x+y)g(x+y)\nu_{t_0}(x, dy) \\
&\leq -\nu_{t_0}(x, (a,\infty) \times (-\infty, -a))
\end{aligned}$$

which implies $\nu_{t_0}(x, (a,\infty) \times (-\infty, -a)) \leq 0$, which means $\nu_{t_0}(x, (a,\infty) \times (-\infty, -a)) = 0$, a contradiction. □

4.2.1 Temporal association

From the previous chapter, we know that for time-homogeneous, stochastically monotone Markov processes, spatial association is equivalent to temporal association if the distribution μ of initial random vector X_0 is in $\mathcal{M}_a$, the space of associated measures (see Theorem 3.7). We would like that this same idea translates in the time-inhomogeneous case, i.e. if X is a time-inhomogeneous Markov process that is stochastically monotone (in the sense of Definition 4.3) and spatially associated (in the sense of Definition 4.1), then X will also be temporally associated (in the sense of Definition 4.2). Lindqvist in [33] proved this notion is true for the case of time-inhomogeneous Markov chains $(X_n)_{n\in\mathbb{N}_0}$ on a finite partially ordered state space E. The author has not seen a proof of an extension of Lindqvist's result

to the continuous-time case with more general state space, like $\mathbb{R}^d$ or a Polish space. But we strongly believe such a result will hold true. We leave that as an exercise to the reader in the following conjecture:

Conjecture 4.9. Let $X = (X_t)_{t\geq 0}$ be a be a time-inhomogeneous, stochastically monotone Markov process. If $X_0 \sim \mu \in \mathcal{M}_a$, and X is spatially associated, then X will be temporally associated.

4.2.2 Other forms of dependence in time-inhomogeneous Markov processes

In the previous chapter, we showed that the Lévy measure condition (3.29) was not only equivalent to spatial association for stochastically monotone jump processes, but also to spatial PUOD, PLOD, POD, PSD, PSA, and WA. These other forms of dependence can analogously be characterized in the time-inhomogeneous setting for the jump processes considered in Theorem 4.8. To do this, we show that (4.40) is a necessary condition for spatial PUOD.

Definition 4.4. Let $X = (X_t)_{t\geq 0}$ be a time-inhomogeneous Markov process on $\mathbb{R}^d$. We say X is **spatially PUOD** if for every $s \leq t$, $x \in \mathbb{R}^d$,

$$\mathbb{E}\left(\prod_{i=1}^{d} f_i(X_t^{(i)}) \,|X_s = x\right) \geq \prod_{i=1}^{d} \mathbb{E}(f_i(X_t^{(i)})|X_s = x), \tag{4.41}$$

where $f_i : \mathbb{R} \to \mathbb{R}_+$ are non-decreasing.

One can alternatively define spatial PUOD of time-inhomogeneous Markov processes by the function class $\mathcal{F}_{ipr}$. For every X_t, let $X_t^{\perp}$ be a random vector with independent components $X_t^{(i),\perp}$, and $X_t^{(i),\perp} \overset{d}{=} X_t^{(i)}$ for all i. X is spatially PUOD if

$$\mathbb{E}(f(X_t)|X_s = x) \geq \mathbb{E}(f(X_t^{\perp})|X_s = x), \quad \forall f \in \mathcal{F}_{ipr},\ s \leq t,\ x \in \mathbb{R}^d. \tag{4.42}$$

We can also use (4.42) to define other forms of spatial dependence in time-inhomogeneous Markov processes, such as PSD, POD, and PLOD, by replacing $\mathcal{F}_{ipr}$ with other function classes, such as $\mathcal{F}_{ism}, \mathcal{F}_{ipr} \cup \mathcal{F}_{dpr}$, and $\mathcal{F}_{dpr}$, respectively.

Theorem 4.10 (Tu, 2017b, [51]). *Let $X = (X_t)_{t\geq 0}$ be a strongly continuous, time-inhomogeneous Markov process with Feller evolution system $(T_{s,t})_{s\leq t}$, generators $(\mathcal{A}_s)_{s\geq 0}$ with rich domains, and that $C_c^\infty(\mathbb{R}^d)$ is the core for $\mathcal{A}_s$, for all $s \geq 0$. Let the corresponding symbols $p_s(x,\xi)$ be s-continuous and bounded as in (4.7) and (4.8) with characteristic triplet $(b_s(x), 0, \nu_s(x,dy))$. If X is spatially PUOD, then $\nu_s(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0$, $\forall s \geq 0$, $x \in \mathbb{R}^d$.*

Proof. Let $\tilde{X} = (\tilde{X}_t)_{t\geq 0}$ on $\mathbb{R}_+ \times \mathbb{R}^d$ be transformation of X, as prescribed in Section 4.1.1, which has Feller semigroup $(\tilde{T}_t)_{t\geq 0}$, generator $(\tilde{\mathcal{A}}, \mathcal{D}(\tilde{\mathcal{A}}))$ with rich domain, bounded symbol $\tilde{p}(\tilde{x}, \tilde{\xi})$, characteristics $(\tilde{b}(\tilde{x}), 0, \tilde{\nu}(\tilde{x}, d\tilde{y}))$ and extended generator $I(\tilde{p})$ on $C_b^2(\mathbb{R}_+ \times \mathbb{R}^d)$, as given to us by Theorem 4.6 and Corollary 4.1.3.

Choose $\tilde{x} = (s, x)$. Let $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}_+$ defined by $f(x_0, ..., x_d) = \prod_{i=0}^d f_i(x_i)$, where $f_i : \mathbb{R} \to \mathbb{R}_+$ are non-decreasing, for all i. Then

$$\begin{aligned}
\tilde{\mathbb{E}}^{\tilde{x}} f(\tilde{X}_t^{(0)}, ..., \tilde{X}_t^{(d)}) &= \tilde{\mathbb{E}}^{\tilde{x}} f(\tilde{X}_t) \\
&= \mathbb{E}(f_{s+t}(X_{s+t}) | X_s = x) \\
&= \mathbb{E}\left(f(s+t, X_{s+t}^{(1)}, ..., X_{s+t}^{(d)}) | X_s = x\right) \\
&= \mathbb{E}\left(f_0(s+t) f_1(X_{s+t}^{(1)}) ... f_d(X_{s+t}^{(d)}) | X_s = x\right) \\
&= f_0(s+t) \cdot \mathbb{E}\left(f_1(X_{s+t}^{(1)}) ... f_d(X_{s+t}^{(d)}) | X_s = x\right) \\
&\geq f_0(s+t) \cdot \prod_{i=1}^d \mathbb{E}(f_i(X_{s+t}^{(i)}) | X_s = x) \\
&= \prod_{i=0}^d \tilde{\mathbb{E}}^{\tilde{x}} f_i(\tilde{X}_t^{(i)})
\end{aligned}$$

where we obtain the inequality by spatial PUOD of process X. Thus, the above calculation shows $\tilde{X}_t$ is PUOD for all $t \geq 0$ in $\mathbb{R}_+ \times \mathbb{R}^d$ with respect $\tilde{\mathbb{P}}^{\tilde{x}}$, for all $\tilde{x}$. By Theorem 3.17, we

have that $\tilde{\nu}(\tilde{x}, (\mathbb{R}^{d+1}_+ \cup \mathbb{R}^{d+1}_-)^c) = 0$ for all $\tilde{x} \in \mathbb{R}_+ \times \mathbb{R}^d$. Observe that the set

$$\{0\} \times (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c \subseteq (\mathbb{R}^{d+1}_+ \cup \mathbb{R}^{d+1}_-)^c.$$

Hence, if $\tilde{x} = (s, x)$,

$$\begin{aligned} 0 &= \tilde{\nu}(\tilde{x}, (\mathbb{R}^{d+1}_+ \cup \mathbb{R}^{d+1}_-)^c) \\ &\geq \tilde{\nu}(\tilde{x}, \{0\} \times (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) \\ &= \nu_s(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) \cdot \delta_0(\{0\}) \\ &= \nu_s(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) \end{aligned}$$

which implies $\nu_s(x, (\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0$, completing our result. □

Remark 4.2.1. Theorem 4.10 also holds true if we replace "PUOD" by "PLOD". This can be easily verified by choosing $f_i : \mathbb{R} \to \mathbb{R}_+$ in the proof of Theorem 4.10 to be non-increasing.

Corollary 4.2.1. *Let $X = (X_t)_{t\geq 0}$ be a strongly continuous, time-inhomogeneous Markov process with the same assumptions as Theorem 4.10. Then condition (4.40) is equivalent to X being associated, WA, PSA, PSD, POD, PUOD, PLOD, in space.*

4.3 Applications and examples

We first present in Section 4.3.1 an important example of time-inhomogeneous Markov processes, called additive processes. These are also called time-inhomogeneous Lévy processes. Such processes are useful in financial models, such as stochastic volatility models with jumps (see [13]). In Section 4.3.2, we show an application of the technique of transformation of time-inhomogeneous to time-homogeneous Markov processes in comparison theorems.

4.3.1 Additive processes

A **process with independent increments (PII)** is a stochastic process $X = (X_t)_{t\geq 0}$ on sample space $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0}, \mathbb{P})$ such that X is càdlàg, adapted, with $X_0 = 0$ a.s. and for all

$s \leq t$, $X_t - X_s$ is independent of $\mathcal{F}_s$. These processes and their semimartingale nature are be described in Jacod and Shiryaev [26].

Definition 4.5. If process $X = (X_t)_{t\geq 0}$ on $\mathbb{R}^d$ is an **additive process** if it is a PII and satisfies stochastic continuity, i.e. $\lim_{h \searrow 0} \mathbb{P}(|X_{t+h} - X_t| \geq a) = 0$, for all $a > 0$.

Thus, observe that can obtain additive processes by relaxing "stationary increments" in the definition of a Lévy process. The following is a theorem of Sato that tells us that additive processes still have "infinitely divisible-like" behavior.

Theorem 4.11 (Sato, [44]). *Let $X = (X_t)_{t\geq 0}$ be an additive process on $\mathbb{R}^d$. Then X_t is infinitely divisible for all $t \geq 0$, and $\phi_{X_t}(u) = \exp(p_t(u))$, where*

$$p_t(u) = iu \cdot b_t - \frac{1}{2} u \cdot \Sigma_t u + \int_{\mathbb{R}^d \setminus \{0\}} (e^{iu\cdot y} - 1 - iu \cdot y\chi(y))\nu_t(dy) \tag{4.43}$$

is the symbol, where for all $t \geq 0$, Σ_t is a symmetric positive semi-definite $d \times d$ matrix, ν_t is a Lévy measure, and $b_t \in \mathbb{R}^d$.

Stochastic continuity of X yields continuity in time of characteristics (b_t, Σ_t, ν_t) and of the characteristic exponent p_t.

Theorem 4.12 (Sato, [44]). *For X additive process with characteristics (b_t, Σ_t, ν_t), we have*

- *Positiveness: $b_0 = 0$, $\Sigma_0 = 0$, $\nu_0 = 0$, and for all $s \leq t$, $\Sigma_t - \Sigma_s$ is a positive definite matrix, and $\nu_t(B) \geq \nu_s(B)$ for all $B \in \mathcal{B}(\mathbb{R}^d)$.*
- *Continuity: if $s \to t$, then $\Sigma_s \to \Sigma_t$, $b_s \to b_t$, and $\nu_s(B) \to \nu_t(B)$ for all $B \in \mathcal{B}(\mathbb{R}^d)$ such that $B \subseteq \{x : |x| \geq \epsilon\}$ for some $\epsilon > 0$.*

Corollary 4.3.1. *Let X be an additive process with characteristic exponents p_t. Then $p_t(u)$ is continuous in t for all $u \in \mathbb{R}^d$.*

Additive processes can also be viewed from the perspective of Markov processes. These processes are time-inhomogeneous, spatially-homogeneous Markov processes, with Markov evolution $(T_{s,t})_{s\leq t}$ given by

$$T_{s,t}f(x) = \mathbb{E}(f(X_t)|X_s = x) = \mathbb{E}f(X_t - X_s + x) \tag{4.44}$$

Such Markov evolutions are also strongly continuous on $C_0(\mathbb{R}^d)$.

Theorem 4.13. *Let X be an additive process with Markov evolution $(T_{s,t})_{s\leq t}$ defined by (4.44). Then $(T_{s,t})_{s\leq t}$ is strongly continuous, i.e. $(T_{s,t})_{s\leq t}$ is a Feller evolution system.*

Proof. We want to show that for all (u, v), where $u \leq v$,

$$\lim_{(u,v)\to(s,t)} ||T_{u,v}f - T_{s,t}f||_\infty = 0, \quad \forall f \in C_0(\mathbb{R}^d).$$

Now let $\epsilon > 0$ and $a > 0$, and pick $f \in C_0(\mathbb{R}^d)$. Then by stochastic continuity, there exists $\delta_0 > 0$ such that if $0 < h < \delta_0$, then $\mathbb{P}(|X_{t+h} - X_t| \geq a) < \epsilon/4||f||_\infty$. Also, by uniform continuity of f, $\exists\delta_1 > 0$ such that if $|z| < \delta_1$, then $\sup_{x\in\mathbb{R}^d} |f(x + z) - f(z)| < \epsilon/2$. Choose $a = \delta_1$. So for all $0 < h < \delta_0$, we have

$$\begin{aligned}
||T_{t,t+h}f - f||_\infty &= \sup_{x\in\mathbb{R}^d} |T_{t,t+h}f(x) - f(x)| \\
&= \sup_{x\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x + z)\mu_{X_{t+h}-X_t}(dz) - f(x) \right| \\
&= \sup_{x\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (f(x + z) - f(x))\mu_{X_{t+h}-X_t}(dz) \right| \\
&\leq \int_{\mathbb{R}^d} \sup_{x\in\mathbb{R}^d} |f(x + z) - f(x)|\mu_{X_{t+h}-X_t}(dz) \\
&= \int_{\{|z|\geq a\}} \sup_{x\in\mathbb{R}^d} |f(x + z) - f(x)|\mu_{X_{t+h}-X_t}(dz) \\
&\quad + \int_{\{|z|\leq a\}} \sup_{x\in\mathbb{R}^d} |f(x + z) - f(x)|\mu_{X_{t+h}-X_t}(dz) \\
&\leq 2||f||_\infty \mathbb{P}(|X_{t+h} - X_t| \geq a) + \int_{\{|z|\leq a\}} \sup_{x\in\mathbb{R}^d} |f(x + z) - f(x)|\mu_{X_{t+h}-X_t}(dz) \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

A similar argument shows $||T_{s-h,h}f - f||_\infty \to 0$. This yields strong continuity at $(u, v) = (0, 0)$. From this, we can obtain the left-continuity of $T_{\cdot,t}$ and right-continuity of $T_{s,\cdot}$. By the boundedness of operator $T_{s,t}$,

$$\begin{aligned}
||T_{s-h,t}f - T_{s,t}f||_\infty &= ||T_{s,t}(T_{t,t+h}f - f)||_\infty \\
&\leq ||T_{s,t}|| \cdot ||T_{t,t+h}f - f||_\infty
\end{aligned}$$

$$\leq ||T_{t,t+h}f - f||_\infty$$
$$\to 0, \quad \text{as } h \searrow 0$$

and by the fact that $T_{s,t}f \in C_0(\mathbb{R}^d)$ when $f \in C_0(\mathbb{R}^d)$,

$$\begin{aligned}||T_{s,t+h}f - T_{s,t}f||_\infty &= ||T_{s-h,s}T_{s,t}f - T_{s,t}f||_\infty \\ &=: ||T_{s-h,s}g - g||_\infty, \quad \text{where } g := T_{s,t}f \in C_0(\mathbb{R}^d) \\ &\to 0, \quad \text{as } h \searrow 0.\end{aligned}$$

Finally, we can show right-continuity of $T_{\cdot,t}$ and left-continuity of $T_{s,\cdot}$ using the same method to show $T_{s,t}$ was continuous at $(s,t) = (0,0)$. To see this, let $s < t$, and choose $h > 0$ small such that $s < s + h < t$. Then

$$\begin{aligned}&||T_{s+h,t}f - T_{s,t}f||_\infty \\ &= \sup_{x\in\mathbb{R}^d} |T_{s+h,t}f(x) - T_{s,t}f(x)| \\ &= \sup_{x\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x+z)\mu_{X_t - X_{s+h}}(dz) - \int_{\mathbb{R}^d} f(x+z)\mu_{X_t - X_s}(dz) \right| \\ &= \sup_{x\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x+z)\mu_{X_t - X_{s+h}}(dz) - \int_{\mathbb{R}^d} f(x+z)\mu_{X_{s+h}-X_s} * \mu_{X_t - X_{s+h}}(dz) \right| \\ &= \sup_{x\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x+z)\mu_{X_t - X_{s+h}}(dz) - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x+z+y)\mu_{X_{s+h}-X_s}(dy)\mu_{X_t - X_{s+h}}(dz) \right| \\ &= \sup_{x\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \left(f(x+z) - \int_{\mathbb{R}^d} f(x+z+y)\mu_{X_{s+h}-X_s}(dy) \right) \mu_{X_t - X_{s+h}}(dz) \right| \\ &\leq \int_{\mathbb{R}^d} \sup_{x\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (f(x+z+y) - f(x+z))\mu_{X_{s+h}-X_s}(dy) \right| \mu_{X_t - X_{s+h}}(dz) \\ &\leq \sup_{x,z\in\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (f(x+z+y) - f(x+z))\mu_{X_{s+h}-X_s}(dy) \right|\end{aligned}$$

where we can show this is small using uniform continuity of f and stochastic continuity of X. A similar technique works for left-continuity of $T_{s,\cdot}$. □

Since additive processes are strongly continuous, they fit into the mold of FESs as described in the beginning of this chapter. It is shown in Cont and Tankov [13] that the

generators $\mathcal{A}_s$ of an additive process has the form

$$\mathcal{A}_s f(x) = b_s \cdot \nabla f(x) + \frac{1}{2}\nabla \cdot \Sigma_s \nabla f(x) + \int_{\mathbb{R}^d \setminus \{0\}} (f(x+y) - f(x) - y \cdot f(x)\chi(y))\nu_s(dy) \quad (4.45)$$

for $f \in C_0^2(\mathbb{R}^d)$. Thus the symbol of the operator $\mathcal{A}_s$ coincides with the characteristic exponent $p_s(\xi)$, which is analogous to symbols and characteristic exponents of Lévy processes. Hence, the additive process has an extended generator, which is an integro-differential operator $I(p_s)$ on $C_b^2(\mathbb{R}^d)$ defined by the RHS of (4.45). Therefore, additive processes are strongly-continuous, time-inhomogeneous Markov processes with symbols $p_s(\xi)$ and characteristics (b_s, Σ_s, ν_s) that do not depend on x, i.e. the state space. They can be classified as strongly continuous, time-inhomogeneous Markov processes that are spatially homogeneous.

Moreover, their FES's $(T_{s,t})_{s\leq t}$ are always stochastically monotone: if $x \leq y$ and $f \in B_b(\mathbb{R}^d) \cap \mathcal{F}_i$, then

$$T_{s,t}f(x) = \mathbb{E}f(X_t - X_s + x) \leq \mathbb{E}f(X_t - X_s + y) = T_{s,t}f(y).$$

Hence, we can apply Theorems 4.7 and 4.8 to additive processes! We have the following characterization.

Theorem 4.14 (Tu, 2017b, [51])**.** *Let $X = (X_t)_{t\geq 0}$ be an additive process with symbols $p_s(\xi)$ and characteristic triplets $(b_s, 0, \nu_s)$. Then X is spatially associated if and only if*

$$\nu_s((\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0, \quad \forall s \in \mathbb{Q}_+. \quad (4.46)$$

Proof. Notice that this is a slightly weaker assumption than the statement of Theorem 4.8. This is because in the case of additive processes, $\nu_s((\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0$, $\forall s \in \mathbb{Q}_+$ implies $\nu_s((\mathbb{R}^d_+ \cup \mathbb{R}^d_-)^c) = 0$, $\forall s \in \mathbb{R}_+$. We show this in $d = 2$.

Assume for contradiction that there is $t_0 \in \mathbb{R}_+ \setminus \mathbb{Q}_+$, such that $\nu_{t_0}((\mathbb{R}^2_+ \cup \mathbb{R}^2_-)^c) > 0$. WLOG, say $\nu_{t_0}((0,\infty) \times (-\infty, 0)) > 0$. By continuity of measure, there exists $a > 0$ such that $\nu_{t_0}((a,\infty) \times (-\infty, -a)) > 0$. By Theorem 4.12, since $A = (a,\infty) \times (-\infty, -a)$ is bounded

away from 0, there exists $(t_n)_{n\in\mathbb{N}} \subset \mathbb{Q}_+$ such that $t_n \to t_0$ and

$$\nu_{t_n}((a,\infty)\times(-\infty,-a)) \to \nu_{t_0}((a,\infty)\times(-\infty,-a)) > 0, \quad \text{as} \quad n\to\infty.$$

Therefore, there exists N large such that for all $n \geq N$, $\nu_{t_n}((a,\infty)\times(-\infty,-a)) > 0$, which is a contradiction. Hence, $\nu_t((\mathbb{R}^2_+ \cup \mathbb{R}^2_-)^c) = 0$ for all $t \geq 0$, which is equivalent to X being spatially associated by Theorem 4.8. □

Corollary 4.3.2. *Let $X = (X_t)_{t\geq 0}$ be an additive process with symbols $p_s(\xi)$ and characteristic triplets $(b_s, 0, \nu_s)$. Then X is spatially PUOD (and also PLOD, POD, PSD, PSA, WA) if and only if ν_s satisfies (4.46).*

Proof. The corollary is a direct result of Theorems 4.10 and 4.14 □

4.3.2 Comparison of Markov processes

Comparing stochastic processes is an important area of research that is intimately related to stochastic orders discussed in Chapter 2, and thus related to dependence orders and the dependence structures we study in this thesis. Let $\mathcal{F}$ be a function class as given in Section 2.2.1, such as $\mathcal{F}_i$. For time-homogeneous Markov processes X and Y with semigroups $(S_t)_{t\geq 0}$ and $(T_t)_{t\geq 0}$ and generators $\mathcal{A}$ and $\mathcal{B}$, respectively, we say that Y **dominates** X **wrt** $\mathcal{F}$ if $S_t f \leq T_t f$, for all $t \geq 0$, for all $f \in \mathcal{F}$. Ruschendorf has proven sufficient conditions for stochastic domination based on the generators.

Theorem 4.15 (Ruschendorf, [42])**.** *Let X, Y be time-homogeneous Markov processes with semigroups $(S_t)_{t\geq 0}$ and $(T_t)_{t\geq 0}$ and generators $\mathcal{A}$ and $\mathcal{B}$. Let $\mathcal{F}$ be one of the class of functions given in Section 2.2.1, and assume $\mathcal{F} \cap C_b(\mathbb{R}^d) \subset \mathcal{D}(\mathcal{A}) \cap \mathcal{D}(\mathcal{A})$. If*

$$S_t f \in \mathcal{F} \text{ for } f \in \mathcal{F} \quad \text{and} \quad \mathcal{A}f \leq \mathcal{B}f \text{ for all } f \in \mathcal{F},$$

then $S_t f \leq T_t f$ for all $f \in \mathcal{F}$.

Proof. See [42]. □

Recent results have given similar results for time-inhomogeneous Markov processes. For time-inhomogeneous Markov processes X and Y with Markov evolutions $(S_{s,t})_{s\leq t}$ and

$(T_{s,t})_{s\leq t}$ and generators $\mathcal{A}_s$ and $\mathcal{B}_s$, respectively, we say that Y **dominates** X **wrt** $\mathcal{F}$ if $S_{s,t}f \leq T_{s,t}f$, for all $s \leq t$, for all $f \in \mathcal{F}$.

Theorem 4.16 (Ruschendorf, [41])**.** *Let X and Y be time-inhomogeneous Markov processes with Markov evolutions $(S_{s,t})_{s\leq t}$ and $(T_{s,t})_{s\leq t}$ and generators $\mathcal{A}_s$ and $\mathcal{B}_s$, respectively. Assume $\mathcal{F} \subset \mathcal{D}(\mathcal{A}_s) \cap \mathcal{D}(\mathcal{B}_s)$. If*

- *$s \mapsto T_{s,t}f$ is right-differentiable for $f \in \mathcal{F}$ for all $0 < s < t$,*
- *$S_{s,t}f \in \mathcal{F}$ for $f \in \mathcal{F}$ for all $s \leq t$,*
- *$\mathcal{A}_s f \leq \mathcal{B}_s f$ for $f \in \mathcal{F}$ for all $s \geq 0$,*

then $S_{s,t}f \leq T_{s,t}f$ for all $f \in \mathcal{F}$, for all $s \leq t$.

Proof. See [41]. □

Remark 4.3.1. The previous result in the time-inhomogeneous case is proven using the a time-inhomogeneous Cauchy problem (see Theorem 3.1 of [41]). Our aim in this section is to provide an alternative proof of Theorem 4.16 using our technique of transforming the time-inhomogeneous process to a time-homogeneous process. From there, we can apply comparison results about homogeneous Markov processes.

First note that if we consider the function class $\mathcal{F}_i$, and assume X and Y are rich Feller processes, then we can extend Theorem 4.15 to integro-differential operators.

Theorem 4.17. *If X and Y are rich Feller processes and have symbols p^X and p^Y, respectively, then if $S_t f \in \mathcal{F}_i$ for $f \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i$ and $I(p^X)f \leq I(p^Y)f$ for all $f \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$, then $S_t f \leq T_t f$ for all $f \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i$.*

Proof. Pick $f \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$. Define $F : [0,\infty) \to C_b(\mathbb{R}^d)$ and $G : [0,\infty) \to C_b(\mathbb{R}^d)$ by

$$F(t) := T_t f - S_t f \tag{4.47}$$

and

$$G(t) := F'(t) - I(p^Y)F(t) = (I(p^Y) - I(p^X))S_t f. \tag{4.48}$$

$G(t) \geq 0$, since $S_t f \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$ and by our assumption. Thus since F, G are continuous (wrt locally uniform convergence), then by Theorem 3.4.5,

$$F(t) = T_t F(0) + \int_0^t T_{t-r} G(r) dr = \int_0^t T_{t-r} G(r) dr \geq 0,$$

giving us our desired result. □

Now we can apply this to prove a version of Ruschendorf's result in Theorem 4.16.

Theorem 4.18. *Let X and Y be strongly-continuous, time-inhomogeneous Markov processes with FESs $(S_{s,t})_{s \leq t}$ and $(T_{s,t})_{s \leq t}$, generators $(\mathcal{A}_s)_{s \geq 0}$ and $(\mathcal{B}_s)_{s \geq 0}$ with rich domains, symbols $p_s(x, \xi)$ and $q_s(x, \xi)$ that are s-continuous and bounded as in (4.7) and (4.8), respectively. Let $C_c^\infty(\mathbb{R}^d)$ be a core for the domains of the generators. Then if X is stochastically monotone (wrt $\mathcal{F}_i$), and*

$$I(p_s) f \leq I(q_s) f, \qquad \text{for all } f \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i, \text{ for all } s \geq 0,$$

then $S_{s,t} f \leq T_{s,t} f$ for all $f \in C_b(\mathbb{R}^d) \cap \mathcal{F}_i$, for all $s \leq t$.

Proof. Let $\tilde{X} = (\tilde{X}_t)_{t \geq 0}$ and $\tilde{Y} = (\tilde{Y}_t)_{t \geq 0}$ on $\mathbb{R}_+ \times \mathbb{R}^d$ be transformations of X and Y, as prescribed in Section 4.1.1, which have Feller semigroups $(\tilde{S}_t)_{t \geq 0}$ and $(\tilde{T}_t)_{t \geq 0}$, generators $(\tilde{\mathcal{A}}, \mathcal{D}(\tilde{\mathcal{A}}))$ and $(\tilde{\mathcal{B}}, \mathcal{D}(\tilde{\mathcal{B}}))$ with rich domain, bounded symbols $\tilde{p}(\tilde{x}, \tilde{\xi})$ and $\tilde{q}(\tilde{x}, \tilde{\xi})$, and extended generators $I(\tilde{p})$ and $I(\tilde{q})$ on $C_b^2(\mathbb{R}_+ \times \mathbb{R}^d)$, respectively, as given to us by Theorem 4.6 and Corollary 4.1.3.

Observe that for all $f \in C_b^2(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{F}_i$, we have

$$\begin{aligned} I(\tilde{p}) f(\tilde{x}) &= \frac{\partial}{\partial s} f(s, x) + I(p_s) f_s(x) \\ &\leq \frac{\partial}{\partial s} f(s, x) + I(q_s) f_s(x) \\ &= I(\tilde{q}) f(\tilde{x}) \end{aligned} \tag{4.49}$$

Now let $h \in C_b^2(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{F}_i$. Fix $s \geq 0$. Then $\tilde{S}_t h|_{\{s\} \times \mathbb{R}^d} \in C_b^2(\{s\} \times \mathbb{R}^d) \cap \mathcal{F}_i$ since X is stochastically monotone. Then there exists $v \in C_b^2(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{F}_i$ that is constant wrt first argument in $\mathbb{R}_+$ and $v(s, x) = \tilde{S}_t h|_{\{s\} \times \mathbb{R}^d}(s, x)$, for all $x \in \mathbb{R}^d$. Then by (4.49), $I(\tilde{p}) v \leq I(\tilde{q}) v$.

This implies that on $\tilde{x} = (s, x)$,

$$I(\tilde{p})(\tilde{S}_t h|_{\{s\}\times\mathbb{R}^d})(s,x) \leq I(\tilde{q})(\tilde{S}_t h|_{\{s\}\times\mathbb{R}^d})(s,x) \tag{4.50}$$

Define $F, G : [0,\infty) \to C_b(\mathbb{R}_+ \times \mathbb{R}^d)$ be defined by

$$F(t) := \tilde{T}_t h - \tilde{S}_t h \tag{4.51}$$

and

$$G(t) := F'(t) - I(\tilde{q})F(t) = (I(\tilde{q}) - I(\tilde{p}))\tilde{S}_t h \tag{4.52}$$

which are both continuous with respect to locally uniform convergence. Then by Theorem 3.4.5, $F(t) = \int_0^t \tilde{T}_{t-r} G(r) dr$. Hence, on $\tilde{x} = (s, x)$,

$$G(r)(\tilde{x}) = (I(\tilde{q}) - I(\tilde{p}))\tilde{S}_t h|_{\{s\}\times\mathbb{R}^d}(s,x) \geq 0$$

by (4.50). Thus, $F(t)(\tilde{x}) \geq 0$. This implies $S_{s,s+t}h_{s+t}(x) \leq T_{s,s+t}h_{s+t}(x)$ for all $x \in \mathbb{R}^d$. Let $f \in C_b^2(\mathbb{R}^d) \cap \mathcal{F}_i$. Then choose $h \in C_b^2(\mathbb{R}_+ \times \mathbb{R}^d) \cap \mathcal{F}_i$ that is constant in the first argument, and $h(\tilde{x}) = f(x)$. Then we have

$$S_{s,s+t}f(x) = S_{s,s+t}h_{s+t}(x) \leq T_{s,s+t}h_{s+t}(x) = T_{s,s+t}f(x),$$

giving us our desired result. □

Chapter 5

Negative dependence in ID distributions and Lévy processes

Analogous to Chapters 2 and 3, we can study various forms of negative dependence in multivariate distributions and stochastic processes. In this chapter, we give an overview of different forms of negative dependence. For general Poissonian distributions, conditions for negative dependence based on the Lévy measure have not been found in the literature by the author. We aim to provide such conditions in this chapter and then extend our results to jump-Lévy processes. Our standard reference on the topic of various forms of negative dependence is [34].

5.1 Various forms of negative dependence

For a multivariate distribution $X = (X_1, ..., X_d)$, we say X is **negatively correlated (NC)** if $\text{Cov}(X_i, X_j) \leq 0$ for all $i \neq j$. For similar reasons to the case of positive correlation (see Chapter 2, one may consider stronger forms of negative dependence. The classical strong form of negative dependence to consider is negative association. To define negative association, one cannot simply take the definition of association in Definition 2.1 and flip the inequality, i.e. say X is negatively associated if $\text{Cov}(f(X), g(X)) \leq 0$ for all $f, g \in \mathcal{F}_i$. No non-trivial random vector would satisfy this, since, for $f = g$, the $\text{Cov}(f(X), f(X)) = Var(f(X)) > 0$, if X and f are non-trivial. Therefore, to define negative association, one

can consider flipping the inequality in the definition of weak association (WA) (Definition 2.2).

Definition 5.1. A random vector $X = (X_1, ..., X_d)$ is **negatively associated (NA)** if, for any pair of disjoint subsets $I, J \subseteq \{1, .., d\}$, with $|I| = k$, $|J| = n$,

$$\text{Cov}(f(X_I), g(X_J)) \leq 0, \tag{5.1}$$

where $X_I := (X_i : i \in I)$, $X_J := (X_j : j \in J)$, for any $f : \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$ non-decreasing,

This definition was first presented by Alam and Saxena [1]. It is easy to see that NA is stronger than negative correlation.

Claim 5.1.1. *If $X = (X_1, ..., X_d)$ is negatively associated, then X is negatively correlated.*

Proof. Choose $i, j \in \{1, ..., d\}$, where $i \neq j$. Choose $I = \{i\}$, $J = \{j\}$. Then I, J disjoint, and now choose $f, g : \mathbb{R} \to \mathbb{R}$ to be the identity functions. By NA,

$$\begin{aligned} 0 &\geq \text{Cov}(f(X_I), g(X_J)) \\ &= \text{Cov}(f(X_i), g(X_j)) \\ &= \text{Cov}(X_i, X_j) \end{aligned}$$

□

We list some useful properties of negative association.

Proposition 5.1.1.

(i) Negative association is preserved under weak convergence.

(ii) Let $X = (X_1, ..., X_d)$ and Y both be NA, and X and Y are independent. Then $X + Y$ is also NA.

(iii) Let $X = (X_1, X_2)$ be weakly associated. Then the random vector $(X_1, -X_2)$ is negatively associated.

(iv) Let $X = (0, ..., 0, X_i, 0, ..., 0, X_j, 0, ..., 0)$ *be a weakly associated random vector in* $\mathbb{R}^d$. *Then the random vector* $X^* := (0, ..., 0, X_i, 0, ..., 0, -X_j, 0, ..., 0)$ *is negatively associated.*

(v) If $X = (X_1, ..., X_d)$ *has independent components, then* X *is NA.*

Proof. The proof of (i) is similar to the proof that weak association is preserved under weak convergence, so we omit that proof here and refer the reader to Chapter 2.

(ii) Let $I := \{s_1, ..., s_k\}, J := \{t_1, ..., t_n\} \subseteq \{1, ..., d\}$ be disjoint sets. Choose $f : \mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ non-decreasing. Define $Z := X + Y$. By independence of X and Y, we have

$$\begin{aligned}\mathbb{E}[f(Z_I)g(Z_J)] &= \mathbb{E}[f(X_{s_1} + Y_{s_1}, ..., X_{s_k} + Y_{s_k})g(X_{t_1} + Y_{t_1}, ..., X_{t_n} + Y_{t_n})] \\ &= \mathbb{E}_X[\mathbb{E}_Y[f(X_{s_1} + Y_{s_1}, ..., X_{s_k} + Y_{s_k})g(X_{t_1} + Y_{t_1}, ..., X_{t_n} + Y_{t_n})]]\end{aligned}$$

We know for all fixed $x_{s_1}, ..., x_{s_k}, x_{t_1}, ..., x_{t_n} \in \mathbb{R}$,

$$\begin{aligned}&\mathbb{E}_Y[f(x_{s_1} + Y_{s_1}, ..., x_{s_k} + Y_{s_k})g(x_{t_1} + Y_{t_1}, ..., x_{t_n} + Y_{t_n})]] \\ &\leq \mathbb{E}_Y[f(x_{s_1} + Y_{s_1}, ..., x_{s_k} + Y_{s_k})]\mathbb{E}_Y[g(x_{t_1} + Y_{t_1}, ..., x_{t_n} + Y_{t_n})]\end{aligned}$$

by NA of Y, since f, g composed with a shift is still non-decreasing. This implies

$$\begin{aligned}&\mathbb{E}_Y[f(X_{s_1} + Y_{s_1}, ..., X_{s_k} + Y_{s_k})g(X_{t_1} + Y_{t_1}, ..., X_{t_n} + Y_{t_n})]] \\ &\leq \mathbb{E}_Y[f(X_{s_1} + Y_{s_1}, ..., X_{s_k} + Y_{s_k})]\mathbb{E}_Y g(X_{t_1} + Y_{t_1}, ..., X_{t_n} + Y_{t_n})], \ \ a.s.\end{aligned}$$

Define $h : \mathbb{R}^k \to \mathbb{R}$ by

$$h(x_{s_1}, ..., x_{s_k}) := \mathbb{E}_Y[f(x_{s_1} + Y_{s_1}, ..., x_{s_k} + Y_{s_k})]$$

and $l : \mathbb{R}^n \to \mathbb{R}$ by

$$l(x_{t_1}, ..., x_{t_n}) = \mathbb{E}_Y[g(x_{t_1} + Y_{t_1}, ..., x_{t_n} + Y_{t_n})].$$

Then by monotonicity of f, g, and the expectation, h, l are also non-decreasing. Therefore,

$$\begin{aligned}
&\mathbb{E}_X[\mathbb{E}_Y[f(X_{s_1}+Y_{s_1},...,X_{s_k}+Y_{s_k})g(X_{t_1}+Y_{t_1},...,X_{t_n}+Y_{t_n})]] \\
&\leq \mathbb{E}_X[\mathbb{E}_Y[f(X_{s_1}+Y_{s_1},...,X_{s_k}+Y_{s_k})]\mathbb{E}_Y[g(X_{t_1}+Y_{t_1},...,X_{t_n}+Y_{t_n})]] \\
&= \mathbb{E}_X[h(X_{s_1},...,X_{s_k})l(X_{t_1},...,X_{t_n})] \\
&\leq \mathbb{E}_X h(X_{s_1},...,X_{s_k})\mathbb{E}_X l(X_{t_1},...,X_{t_n}) \\
&= \mathbb{E}_X[\mathbb{E}_Y[f(X_{s_1}+Y_{s_1},...,X_{s_k}+Y_{s_k})]] \cdot \mathbb{E}_X[\mathbb{E}_Y[g(X_{t_1}+Y_{t_1},...,X_{t_n}+Y_{t_n})]] \\
&= \mathbb{E}f(Z_I)\mathbb{E}g(Z_J)
\end{aligned}$$

which gives us our desired result.

(iii) Let $f, g : \mathbb{R} \to \mathbb{R}$ be non-decreasing. We want to show $\mathrm{Cov}(f(X_1), g(-X_2)) \leq 0$. Know, for all $h, k : \mathbb{R} \to \mathbb{R}$ non-decreasing, $\mathrm{Cov}(h(X_1), k(X_2)) \geq 0$ by WA of X. Equivalently, we also have, for all $h, k : \mathbb{R} \to \mathbb{R}$ *non-increasing*, $\mathrm{Cov}(h(X_1), k(X_2)) \geq 0$, i.e. weak association is true for all pairs of non-increasing functions. Define $h(x) := f(x)$, where f is non-decreasing, and $k(y) := g(-y)$, where g is non-decreasing. Therefore, $-h(x) = -f(x)$ and $k(y) = g(-y)$ are non-increasing functions. By weak association of (X_1, X_2), we have

$$\begin{aligned}
0 &\leq \mathrm{Cov}(-h(X_1), k(X_2)) \\
&= \mathbb{E}[-h(X_1)k(X_2)] - \mathbb{E}[-h(X_1)]\mathbb{E}[k(X_2)] \\
&= \mathbb{E}[-f(X_1)g(-X_2)] - \mathbb{E}[-f(X_1)]\mathbb{E}[g(-X_2)] \\
&= -\mathbb{E}[f(X_1)g(-X_2)] + \mathbb{E}[f(X_1)]\mathbb{E}[g(-X_2)] \\
&= -\mathrm{Cov}(f(X_1), g(-X_2))
\end{aligned}$$

which implies $\mathrm{Cov}(f(X_1), g(-X_2)) \leq 0$.

(iv) Let $I, J \subset \{1, ..., d\}$ be disjoint subsets such that $|I| = k$ and $|J| = n$. WLOG, say $i \in I$ and $j \in J$. We want to show $\text{Cov}(f(X_I^*), g(X_J^*) \leq 0$ for all $f : \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$ non-decreasing. Define $v : \mathbb{R}^k \to \mathbb{R}$ by $v = -f$ and $w : \mathbb{R}^n \to \mathbb{R}$ by $w(x) = g(-x)$. Then v, w are non-increasing functions. Therefore, by WA,

$$\begin{aligned}
0 &\leq \text{Cov}(v(X_I), w(X_J)) \\
&= \mathbb{E}(-f(X_I)g(-X_J)) - \mathbb{E}(-f(X_I)) \cdot \mathbb{E}g(-X_J) \\
&= -\mathbb{E}(f(X_I)g(-X_J)) + \mathbb{E}f(X_I) \cdot \mathbb{E}g(-X_J) \\
&= -\mathbb{E}(f(0, ..., 0, X_i, 0, ..., 0)g(0, ..., 0, -X_j, 0, ..., 0)) \\
&\quad + \mathbb{E}f(0, ..., 0, X_i, 0, ..., 0) \cdot \mathbb{E}g(0, ..., 0, -X_j, 0, ..., 0) \\
&= -\text{Cov}(f(X_I^*), g(X_J^*)),
\end{aligned}$$

which implies $\text{Cov}(f(X_I^*), g(X_J^*)) \leq 0$.

(v) This proof is similar to the case of "independent components" implies WA.

□

We can also define various forms of negative dependence based on stochastic orderings like how we did in Chapter 3. Definitions 5.2, 5.3, and 5.4 were first presented by Lehmann [31] in $d = 2$. Let $X = (X_1, ..., X_d)$, and let $X^\perp = (X_1^\perp, ..., X_d^\perp)$ have independent components and $X_i^\perp \overset{d}{=} X_i$ for all $i = 1, ..., d$.

Definition 5.2. X is **negative upper orthant dependent (NUOD)** if $X \leq_{uo} X^\perp$. Equivalently, X satisfies, for all $t = (t_1, ..., t_d) \in \mathbb{R}^d$,

$$\mathbb{P}(X_1 > t_1, ..., X_d > t_d) \leq \mathbb{P}(X_1 > t_1)...\mathbb{P}(X_d > t_d). \tag{5.2}$$

Definition 5.3. X is **negative lower orthant dependent (NLOD)** if $X \leq_{lo} X^\perp$. Equivalently, X satisfies, for all $t = (t_1, ..., t_d) \in \mathbb{R}^d$,

$$\mathbb{P}(X_1 \leq t_1, ..., X_d \leq t_d) \leq \mathbb{P}(X_1 \leq t_1)...\mathbb{P}(X_d \leq t_d). \tag{5.3}$$

Definition 5.4. X is **negative orthant dependent (NOD)** if $X \leq_c X^\perp$, i.e. if X is NUOD and NLOD, or equivalently, satisfying (5.2) and (5.3).

Definition 5.5. X is **negative supermodular dependent (NSD)** if $X \leq_{sm} X^\perp$

Many of the same properties that hold for PUOD, PLOD, POD, and PSD also hold for their "negative" counterparts. We list those properties in the following proposition.

Proposition 5.1.2.

(i) *X is NUOD (NLOD) if and only if $X \leq_{\mathcal{F}_{ipr}} X^\perp$ ($\leq_{\mathcal{F}_{dpr}}$).*

(ii) *X is NOD if and only if $X \leq_{\mathcal{F}_{ipr}} X^\perp$ and $X \leq_{\mathcal{F}_{dpr}} X^\perp$.*

(iii) *Let X be NOD (NUOD, NLOD, respectively) and $\Phi = (\Phi_1, ..., \Phi_k) : \mathbb{R}^d \to \mathbb{R}^k$ be defined as $\Phi(x_1, ..., x_d) = (\Phi_1(x_1), ..., \Phi_k(x_k))$, where $k \leq d$, and Φ_i's non-decreasing on $\mathbb{R}$. Then $\Phi(X)$ is NOD (NUOD, NLOD, respectively).*

(iv) *Let X be NSD. Then for any $g_1, ..., g_d : \mathbb{R} \to \mathbb{R}$ non-decreasing, $(g_1(X_1), ..., g_d(X_d))$ is also NSD.*

(v) *Let X be NOD (NUOD, NLOD, respectively). For all $I \subseteq \{1, ..., d\}$, we have that $X_I = (X_i : i \in I)$ is NOD (NUOD, NLOD, respectively).*

Proof. Given the similarity of the proofs of the above statements with the proofs of the "positive" version of these statements for POD, PUOD, PLOD, and PSD, we omit the proof here. Instead, we refer the reader to Chapter 3 and to [34]. □

The strengths of the different negative dependence forms given so far are analogous to the strengths in the positive dependence case. Figure 5.1 and the following proposition outline the different levels of strength.

Proposition 5.1.3.

(i) *(NA) implies (NSD)*

(ii) *(NSD) implies (NOD)*

(iii) *(NOD) implies (NUOD)*

(iv) *(NOD) implies (NLOD)*

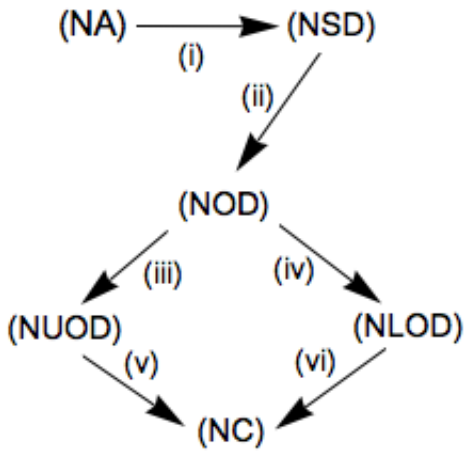

Figure 5.1: Implication map of negative dependence

(v) (NUOD) implies (NC) *(vi) (NLOD) implies (NC)*

Proof. Let $X = (X_1, ..., X_d)$ be a random vector and $X^\perp = (X_1^\perp, ..., X_d^\perp)$, where components of $X^\perp$ are independent, and $X_i^\perp \overset{d}{=} X_i$ for all $i \in \{1, ..., d\}$.

(i) For this proof, we refer the reader to [12].

(ii) We know, by NSD, f supermodular implies $\mathbb{E}f(X^\perp) \geq \mathbb{E}f(X)$. Fix $t = (t_1, ..., t_d)$. Define $f(x_1, ..., x_d) := \mathbb{1}_{(t_1,\infty)}(x_1)...\mathbb{1}_{(t_d,\infty)}(x_d)$. This function is supermodular by Lemma 2.2.2, since $\mathbb{1}_{(t_i,\infty)}(x_i)$ is non-decreasing and non-negative for all i. Hence, by NSD,

$$\begin{aligned}
\mathbb{P}(X_1 > t_1)...\mathbb{P}(X_d > t_d) &= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1)]...\mathbb{E}[\mathbb{1}_{(t_d,\infty)}(X_d)] \\
&= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1^\perp)]...\mathbb{E}[\mathbb{1}_{(t_d,\infty)}(X_d^\perp)] \\
&= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1^\perp)...\mathbb{1}_{(t_d,\infty)}(X_d^\perp)], \text{ by independence} \\
&= \mathbb{E}[f(X^\perp)] \\
&\geq \mathbb{E}[f(X)] \\
&= \mathbb{E}[\mathbb{1}_{(t_1,\infty)}(X_1)...\mathbb{1}_{(t_d,\infty)}(X_d)] \\
&= \mathbb{P}(X_1 > t_1, ..., X_d > t_d),
\end{aligned}$$

giving us our desired result. We can do a similar trick to show the negative lower orthant dependence by considering f_i in Lemma 2.2.2 to be all non-increasing.

(iii) True by definition of NOD.

(iv) True by definition of NOD.

(v) Assume $\text{Cov}(X_i, X_j) < \infty$ for all i, j. If X is NUOD, then every bivariate vector (X_i, X_j) is NUOD by Proposition 5.1.2(v). Thus,

$$\begin{aligned}\text{Cov}(X_i, X_j) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \text{Cov}(\mathbb{1}_{(t,\infty)}(X_i), \mathbb{1}_{(s,\infty)}(X_j)) ds\, dt \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} [\mathbb{P}(X_i > t, X_j > s) - \mathbb{P}(X_i > t)\mathbb{P}(X_j > s)] ds\, dt \\ &\leq 0\end{aligned}$$

since the integrand is less than or equal to 0 by (X_i, X_j) NUOD.

(vi) Assume $\text{Cov}(X_i, X_j) < \infty$ for all i, j. If X is NLOD, then every bivariate vector (X_i, X_j) is NLOD by Proposition 5.1.2(v). Additionally, (X_i, X_j) is NLOD if and only if (X_i, X_j) is NUOD, since NLOD and NUOD are equivalent notions in dimension 2 (see [34]). Hence, $\text{Cov}(X_i, X_j) \leq 0$ by the same argument as above.

□

Again, as in the case of positive dependence, the above dependencies of NA, NSD, NOD, NUOD, NLOD, and NC are not equivalent in general. We can generate examples similar to the ones used in Chapter 2 to show that the different forms of positive dependence are not equivalent. Instead of showing all the different examples here, we leave this as an exercise to the reader or refer the reader to [34] and [9]. We do note a special case where they become equivalent.

Proposition 5.1.4. *If $X = (X_1, X_2)$ is a random vector in $\mathbb{R}^d$, then the following negative dependence structures are equivalent: NA, NSD, NOD, NUOD, NLOD.*

Proof. The proof of this is similar to the proof of Proposition 2.2.5 in Chapter 2. □

The case of Lévy processes will also yield a scenario where these negative dependence structures are equivalent. This will be discussed in Section 5.3

5.2 Negative dependence and infinite divisibility

Some of the results presented in this section and Section 5.3 will be a part of a joint paper with Jan Rosinski [40] that will treat association and negative association of infinitely divisible distributions and processes and Lévy processes under a unified framework.

For Gaussian random vectors, negative association has been characterized.

Theorem 5.1 (Joag-Dev, Proschan [27]). *Let $X \sim N(b, \Sigma)$ be a Gaussian random vector in $\mathbb{R}^d$. Then X is negatively associated if and only if $\Sigma \leq 0$, i.e. is negatively correlated.*

Little has been done on Poissonian random vectors from what the author has seen. Work has been done on characterizing negative association for α-stable random vectors [30], but not general Poissonian distributions. Our aim here is to provide a sufficient condition for negative association of Poissonian random vectors based on the Lévy measure.

Theorem 5.2 (Lee et. al., [30]). *Let $X = (X_1, ..., X_d)$ be an α-stable random vector with spectral measure ρ. Then X is NA if and only if*

$$\rho(\{(s_1, ..., s_d) \in S^d : \text{ for some } i \neq j, s_i s_j > 0\}) = 0.$$

Given the close connection between the spectral measure and the Lévy measure of X α-stable (see (2.23)), we can use Theorem 5.2 to motivate a sufficient condition for NA of Poissonian distributions.

Claim 5.2.1. *Let $X = (X_1, ..., X_d) \sim ID(b, 0, \nu)$. If*

$$\nu(\{x \in \mathbb{R}^d : \text{ for some } i \neq j, x_i x_j > 0\}) = 0, \tag{5.4}$$

then X is negatively associated.

We first present a cute proof in $d = 2$. Then we generalize to higher dimensions.

Lemma 5.2.1. *Let N be a Poisson random measure with intensity ν on $(\mathbb{R}^2 \setminus \{0\}, \mathcal{B}(\mathbb{R}^2 \setminus \{0\}))$. Define $\mathbb{R}^2_{+,-} := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0,\ x_2 \leq 0\}$ and $\mathbb{R}^2_{-,+} := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \leq 0,\ x_2 \geq 0\}$. If $g : \mathbb{R}^2 \to \mathbb{R}^2_{+,-}$, and $\int_{\mathbb{R}^2} g_i(x)N(dx) < \infty$ for $i = 1, 2$, then $\int_{\mathbb{R}^2} g(x)N(dx)$ is negatively associated. Similarly, if $h : \mathbb{R}^2 \to \mathbb{R}^2_{-,+}$, then $\int_{\mathbb{R}^2} h(x)N(dx)$ is NA.*

Proof. Let $g : \mathbb{R}^2 \to \mathbb{R}^2_{+,-}$, where $g = (g_1, g_2)$, with $g_1 : \mathbb{R}^2 \to \mathbb{R}_+$ and $g_2 : \mathbb{R}^2 \to \mathbb{R}_-$. Then $g^* := (g_1, -g_2)$ satisfies $g^* : \mathbb{R}^2 \to \mathbb{R}^2_+$. Therefore, by Lemma 2.3.1,

$$\int_{\mathbb{R}^2} g^*(x)N(dx) = \left(\int_{\mathbb{R}^2} g_1(x)N(dx), -\int_{\mathbb{R}^2} g_2(x)N(dx) \right)$$

is associated. Hence, by Proposition 5.1.1,

$$\int_{\mathbb{R}^2} g(x)N(dx) = \left(\int_{\mathbb{R}^2} g_1(x)N(dx), \int_{\mathbb{R}^2} g_2(x)N(dx) \right)$$

is negatively associated. We can apply a similar technique to show $\int h(x)N(dx)$ is NA. □

Proof of Claim 5.2.1 ($d = 2$).

Proof. In $d = 2$, (5.4) becomes $\nu(\mathbb{R}^2_+ \cup \mathbb{R}^2_-) = 0$. WLOG we can assume $b = 0$. By Lévy-Itô decomposition, we have

$$X = \int_{|x|<1} x\tilde{N}(dx) + \int_{|x|\geq 1} xN(dx), \quad a.s.$$

where N is a Poisson random measure with intensity measure ν. Define $Y := (Y_1, Y_2) := \int_{|x|<1} x\tilde{N}(dx)$ and $Z := (Z_1, Z_2) = \int_{|x|\geq 1} xN(dx)$. Let's first show Z is NA. By assumption of the concentration of ν, we have

$$Z = \int_{|x|\geq 1} xN(dx) = \int_{\{|x|\geq 1\}\cap\mathbb{R}^2_{+,-}} xN(dx) + \int_{\{|x|\geq 1\}\cap\mathbb{R}^2_{-,+}} xN(dx)$$

Define $g : \mathbb{R}^2 \to \mathbb{R}^2_{+,-}$ by

$$g(x) = x\mathbb{1}_{\{|x|\geq 1\}\cap\mathbb{R}^2_{+,-}}(x)$$

and $h : \mathbb{R}^2 \to \mathbb{R}^2_{-,+}$ by

$$h(x) = x\mathbb{1}_{\{|x|\geq 1\}\cap\mathbb{R}^2_{-,+}}(x)$$

By Lemma 5.2.1,

$$\int_{\mathbb{R}^2} g(x)N(dx) = \int_{\{|x|\geq 1\}\cap\mathbb{R}^2_{+,-}} xN(dx) \tag{5.5}$$

and

$$\int_{\mathbb{R}^2} h(x)N(dx) = \int_{\{|x|\geq 1\}\cap\mathbb{R}^2_{-,+}} xN(dx) \tag{5.6}$$

are both NA. Since (5.5) and (5.6) are Poisson integrals over disjoint sets, they are independent, and therefore, by Proposition 5.1.1, $Z = \int_{|x|\geq 1} xN(dx)$ is NA. Now let's show Y is NA. Define $Y_n := \int_{\{\frac{1}{n}<|x|<1\}} x\tilde{N}(dx)$. Observe that

$$\int_{\{\frac{1}{n}<|x|<1\}} xN(dx) = \int_{\{\frac{1}{n}<|x|<1\}\cap\mathbb{R}^2_{+,-}} xN(dx) + \int_{\{\frac{1}{n}<|x|<1\}\cap\mathbb{R}^2_{-,+}} xN(dx)$$

is NA by the same argument for Z to be NA. This implies Y_n is NA, since $\tilde{N}$ just contributes a non-random shift that doesn't affect the dependence. Since $Y_n \xrightarrow{L^2} Y$ and, thus $Y_n \Rightarrow Y$, and NA is preserved under weak convergence, we have that Y is NA. By Proposition 5.1.1, $X = Y + Z$ is NA. □

Now here is the proof of Claim 5.2.1 for general dimension d.

Proof of Claim 5.2.1

Proof. By assumption (5.4), we have that ν is concentrated on the sets

$$A_{i,j} = \{0\} \times ... \times \{0\} \times \mathbb{R}_+ \times \{0\} ... \times \{0\} \times \mathbb{R}_- \times \{0\} \times ... \times \{0\}$$

and

$$B_{i,j} = \{0\} \times ... \times \{0\} \times \mathbb{R}_- \times \{0\} ... \times \{0\} \times \mathbb{R}_+ \times \{0\} \times ... \times \{0\}.$$

where $i \neq j$ and $i, j \in \{1, ..., d\}$, where the $\mathbb{R}_+, \mathbb{R}_-$ are the ith and jth positions, respectively, for $A_{i,j}$ and the reverse for $B_{i,j}$. WLOG, assume $b = 0$. By Lévy-Itô decomposition,

$$X = \int_{\{|x|<1\}} x\tilde{N}(dx) + \int_{\{|x|\geq 1\}} xN(dx) =: Y + Z \tag{5.7}$$

Let's first show Z is NA.

$$Z = \int_{\{|x|\geq 1\}} xN(dx) = \sum_{\substack{i,j \\ i\neq j}} \int_{\{|x|\geq 1\}\cap A_{i,j}} xN(dx) + \sum_{\substack{i,j \\ i\neq j}} \int_{\{|x|\geq 1\}\cap B_{i,j}} xN(dx) \tag{5.8}$$

Define $g(x) = x \cdot \mathbb{1}_{\{|x|\geq 1\}\cap A_{i,j}}$. Then $g : \mathbb{R}^d \to A_{i,j}$. Hence

$$\begin{aligned} \int_{\{|x|\geq 1\}\cap A_{i,j}} xN(dx) &= \int_{\mathbb{R}^d} g(x)N(dx) \\ &= \left(\int_{\mathbb{R}^d} g_1(x)N(dx), ..., \int_{\mathbb{R}^d} g_d(x)N(dx)\right) \\ &= \left(0, ..., 0, \int_{\mathbb{R}^d} g_i(x)N(dx), 0, ..., 0, \int_{\mathbb{R}^d} g_j(x)N(dx), 0, ..., 0\right) \end{aligned} \tag{5.9}$$

Define $g* : \mathbb{R}^d \to \mathbb{R}^d_+$ by

$$g^*(x) = (g_1(x), ..., -g_j(x), ..., g_d(x)) = (0, ..., 0, g_i(x), 0, ..., 0, -g_j(x), 0, ..., 0).$$

Then by Lemma 2.3.1,

$$\int_{\mathbb{R}^d} g^*(x)N(dx) = \left(0, ..., 0, \int_{\mathbb{R}^d} g_i(x)N(dx), 0, ..., 0, -\int_{\mathbb{R}^d} g_j(x)N(dx), 0, ..., 0\right) \tag{5.10}$$

is associated. Hence, by Proposition 5.1.1, $\int_{\mathbb{R}^d} g(x)N(dx)$ is associated. We can use this technique to show that each of the Poisson integrals in (5.8) is NA. Since $A_{i,j}$'s and $B_{i,j}$'s are all disjoint, the Poisson integrals are all independent, and by Proposition 5.1.1, Z is NA. Define for all $n \in \mathbb{N}$,

$$Y_n := \int_{\{1/n<|x|<1\}} x\tilde{N}(dx) \tag{5.11}$$

By the same argument used for random vector Z, we can show that each Y_n is NA. Since NA is preserve under weak convergence, and we have $Y_n \Rightarrow Y$, then Y is also NA. Since Y and Z are independent, X is NA. □

Remark 5.2.1. For $d > 2$, Claim 5.2.1 may be too strong and could miss a collection of Poissonian distributions. For example, if ν is absolutely continuous with respect to the Lebesgue measure λ, then ν could not be concentrated on sets $A_{i,j}$ and $B_{i,j}$ given in the proof of Claim 5.2.1. This is because the Lebesgue measure of such sets is always 0, thus making $\nu = 0$ on those sets. A way around this is the consider changing assumption (5.4) to be assumptions on projections of the Lévy measure.

Theorem 5.3 (Rosinski, Tu, 2017, [40])**.** *Let $X \sim ID(b, 0, \nu)$ in $\mathbb{R}^d$, where $d > 2$. Define $\nu^{ij} := \boldsymbol{proj}_{ij}(\nu)$ to be the projection of Lévy measure ν onto the ith and j^{th} coordinates, $i \neq j$. If for all $i \neq j$,*

$$\nu^{ij}(\{x = (x_i, x_j) \in \mathbb{R}^2 : x_i x_j > 0\}) = \nu^{ij}(\mathbb{R}^2_+ \cup \mathbb{R}^2_-) = 0, \tag{5.12}$$

then X is NA.

Proof. WLOG, assume $X \sim ID(0, 0, \nu)$. Observe that for all $A \in \mathcal{B}(\mathbb{R}^2 \setminus \{0\})$, $\nu^{ij}(A) = \nu(\pi_{ij}^{-1}(A \setminus (0,0))) = \nu(x \in \mathbb{R}^d : \pi_{ij}x \in A \setminus (0,0))$, where π_{ij} is the projection onto the ith and j^{th} coordinates: $\pi_{ij}x = (x_i, x_j)$. By Lévy-Itô decomposition,

$$X = \int_{\{|x|<1\}} x\tilde{N}(dx) + \int_{\{|x|\geq 1\}} xN(dx) =: Y + Z, \tag{5.13}$$

where N is a PRM with intensity ν. We first show Z is NA. Define the following sets for $i \neq j$: $A_{i,j} = \{x \in \mathbb{R}^d : x_i x_j \neq 0\}$, $A'_{i,j} = A_{i,j} \cap \{|x| \geq 1\}$, and for all i: $B_i = \{x \in \mathbb{R}^d : x_i \neq 0, x_j = 0, \ \forall j \neq i\}$, $B'_i = B_i \cap \{|x| \geq 1\}$.

By condition (5.12), we have that

$$\nu(A_{i,j} \cap A_{k,l}) = 0, \text{ for all } (i,j) \neq (k,l). \tag{5.14}$$

To see this, note that it is sufficient to show that for $(i,j) \neq (i,l)$, $\nu(A_{i,j} \cap A_{i,l}) = 0$. The set $A_{i,j} \cap A_{i,l} = \{x \in \mathbb{R}^d : x_i, x_j, x_l \neq 0\}$ can be partitioned into eight disjoint sets of the form

$$C_{i,j,l}^{s_i,s_j,s_l}, \qquad s_i, s_j, s_l \in \{+,-\} \tag{5.15}$$

So if $x \in C_{i,j,l}^{s_i,s_j,s_l}$, the s_i, s_j, s_l determine the sign of the $i^{\text{th}}, j^{\text{th}}, l^{\text{th}}$ component of x. For example,

$$C_{i,j,l}^{+,-,+} = \{x \in \mathbb{R}^d : x_i > 0, x_j < 0, x_l > 0\} \tag{5.16}$$

Thus,

$$A_{i,j} \cap A_{i,l} = \bigcup_{s_i,s_j,s_l \in \{+,-\}} C_{i,j,l}^{s_i,s_j,s_l} \tag{5.17}$$

By (5.12), $\nu(C_{i,j,l}^{s_i,s_j,s_l}) = 0$. For example,

$$\begin{aligned}
\nu(C_{i,j,l}^{+,-,+}) &= \nu(x : x_i > 0, x_j < 0, x_l > 0) \\
&\leq \nu(x : x_i x_l > 0) \\
&= \nu^{il}(\mathbb{R}_+^2) \\
&= 0,
\end{aligned}$$

which implies $\nu(A_{i,j} \cap A_{i,l}) = 0$.

We can decompose Z into a sum of independent random vectors in the following way:

$$\begin{aligned}
Z &= \sum_{i<j} \left(\int_{A'_{i,j}} xN(dx) \right) + \left(\int_{B'_1} x_1 N(dx), ..., \int_{B'_d} x_d N(dx) \right) \\
&= \sum_{i<j} \left(\int_{A'_{i,j}} x_1 N(dx), ..., \int_{A'_{i,j}} x_d N(dx) \right) + \left(\int_{B'_1} x_1 N(dx), ..., \int_{B'_d} x_d N(dx) \right).
\end{aligned} \tag{5.18}$$

Since B_k's are all disjoint, $\{N(B_k)\}_{k=1}^d$ are independent. Also, by condition (5.14), if $k \notin \{i,j\}$, then $\int_{A'_{i,j}} x_k N(dx) = 0$. Therefore, the stochastic integrals on the RHS of (5.18) can be written as:

$$\int_{A'_{i,j}} xN(dx) = \left(0, ..., 0, \int_{A'_{i,j}} x_i N(dx), 0, ..., 0, \int_{A'_{i,j}} x_j N(dx), 0, ..., 0 \right), \tag{5.19}$$

and, hence, (5.14) yields that the stochastic integrals over $N(A_{i,j})$'s are independent. Also, by disjointness of B_k's with $A_{i,j}$'s we also have $\{N(B_k)\}_{k=1}^d$ is independent of $\{N(A_{i,j})\}_{i<j}$. Thus, since all the random vectors on the RHS of (5.18) are independent of each other, it suffices to show that each random vector is negatively associated.

The final random vector on the RHS of (5.18), which is the stochastic integral over the B_k's, has independent components by disjointness of B_k's and is, therefore, by Proposition 5.1.1(v), negatively associated. So now we just need to show each $\int_{A'_{i,j}} xN(dx)$ is NA.

Define $D^+_{i,j} = \mathbb{R} \times ... \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times ... \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times ... \times \mathbb{R}$, where $\mathbb{R}_+$'s are in the ith and j^{th} coordinates. Define similarly $D^-_{i,j}$ to be the same as $D^+_{i,j}$, with $\mathbb{R}_+$ replaced by $\mathbb{R}_-$. Define function $g_+ : \mathbb{R}^d \to \mathbb{R}^d_+$ by

$$g_+(x_1, ..., x_d) = (0, ..., 0, x_i, 0, ..., 0, x_j, 0, ..., 0)\mathbb{1}_{A'_{i,j} \cap D^+_{i,j}} \tag{5.20}$$

and $g_- : \mathbb{R}^d \to \mathbb{R}^d_-$ similarly as g_+, by replacing $D^+_{i,j}$ with $D^-_{i,j}$. Then by condition (5.12),

$$\left(0, ..., 0, \int_{A'_{i,j}} x_i N(dx), 0, ..., 0, -\int_{A'_{i,j}} x_j N(dx), 0, ..., 0\right) = \int_{\mathbb{R}^d} g_+(x)N(dx) + \int_{\mathbb{R}^d} g_-(x)N(dx), \tag{5.21}$$

where the two Poisson integrals on RHS of (5.21) are independent of each other. Then by Lemma 2.3.1, each Poisson integral on RHS of (5.21) is associated. Therefore, the LHS of (5.21) is associated. Hence, Proposition 5.1.1(iv), $\int_{A'_{i,j}} xN(dx)$ is NA for all $i < j$. Thus Z is NA.

We can similarly show that Y is NA, by showing $\int_{\{1/n<|x|<1\}} x\tilde{N}(dx)$ is NA for all n by using a similar technique as above. Then take $n \to \infty$ and use Proposition 5.1.1(i) to show Y is NA. □

The assumption on the projections of the Lévy measure in condition (5.12) also says something interesting about the negative dependence of pairs (X_i, X_j) and the negative association of the entire random vector. Normally, if (X_i, X_j) are negatively associated for all $i \neq j$, then this does *not* imply $X = (X_1, ..., X_d)$ is negatively associated, and a similar notion in the case for other negative dependence structures. But condition (5.12) yields a scenario where this implication is true, according to the result of Theorem 5.3. Moreover,

we know that in $d = 2$, NA is equivalent to PUOD and PLOD. Hence, weaker notions of negative dependence of pairs of components of X become equivalent to negative association of X in $\mathbb{R}^d$.

Corollary 5.2.1. *Let $X = (X_1, ..., X_d) \sim ID(b, 0, \nu)$ such that $\nu^{ij}(\mathbb{R}^2_+ \cup \mathbb{R}^2_-) = 0$ for all $i \neq j$. Then $X = (X_1, ..., X_d)$ is NA if and only if (X_i, X_j) is NUOD (or NLOD) for all $i \neq j$.*

Proof. This is a direct result of Theorem 5.3 and Proposition 5.1.4. □

In general, condition (5.4) is only a sufficient condition. We present an example of an infinitely divisible random vector not satisfying (5.4), but is NA. Our construction is based on the construction of Samorodnitsky in [43].

Theorem 5.4 (Samorodnitsky, [43])**.** *Let Y be an infinitely divisible random variable such that $\mathbb{P}(Y < 0) > 0$, and let $Y_1, Y_2 \overset{iid}{\sim} Y$ such that $Y_1 \leq_{\mathcal{F}_i} Y_1 + Y_2$. Let $N \sim Poiss(1)$, and define independent sequences of independent random variables $(Y_j^{(1)})_{j\in\mathbb{N}}, (Y_j^{(2)})_{j\in\mathbb{N}} \overset{iid}{\sim} Y$, which are both independent of N and Y_1, Y_2. Then*

$$X = (X_1, X_2) = \left(\sum_{j=1}^{N+1} Y_j^{(1)}, \sum_{j=1}^{N+1} Y_j^{(2)}\right) = \left(\sum_{j=1}^{N} Y_j^{(1)}, \sum_{j=1}^{N} Y_j^{(2)}\right) + (Y_1, Y_2) \tag{5.22}$$

is infinitely divisible, associated, and its Lévy measure satisfies

$$\nu(\{x \in \mathbb{R}^2 : x_1 x_2 < 0\}) > 0. \tag{5.23}$$

We construct a random vector X^* in $\mathbb{R}^2$ that is infinitely divisible, NA, and failing condition (5.4).

Theorem 5.5. *There is an infinitely divisible, negatively associated random vector in $\mathbb{R}^2$ such that its Lévy measure ν^* satisfies*

$$\nu^*(\{x \in \mathbb{R}^2 : x_1 x_2 > 0\}) > 0. \tag{5.24}$$

Proof. Take X, N, Y, and the corresponding sequences $(Y_j^{(i)})_{j\in\mathbb{N}}$ from Theorem 5.4. Define $X^* = (X_1, -X_2)$. By Proposition 5.1.1, X^* is NA. Observe that

$$X^* = (X_1, -X_2) = \left(\sum_{j=1}^{N} Y_j^{(1)}, \sum_{j=1}^{N} -Y_j^{(2)}\right) + (Y_1, -Y_2)$$

If we define $Y_j^* = (Y_j^{(1)}, -Y_j^{(2)})$, then $(Y_j^*)_{j\in\mathbb{N}}$ is an iid sequence, which makes

$$\sum_{j=1}^{N} Y_j^* = \left(\sum_{j=1}^{N} Y_j^{(1)}, \sum_{j=1}^{N} -Y_j^{(2)}\right)$$

a compound Poisson random vector, and therefore infinitely divisible. Also, $(Y_1, -Y_2)$ is clearly also infinitely divisible and is independent of $\sum_{j=1}^{N} Y_j^*$, which makes X^* infinitely divisible. Now we just need to show that ν^*, the Lévy measure of X^* satisfies (5.24). Define $Z^* := \sum_{j=1}^{N} Y_j^*$. It is sufficient to show that the Lévy measure of Z^*, ν_{Z^*}, satisfies (5.24), since the Lévy measure of the sum of two independent, infinitely divisible random vectors is just the sum of the corresponding Lévy measures. We know that the Lévy measure of a compound Poisson distribution gives us that $\nu_{Z^*} = \mu^* - \mu^*\{0\}\delta_0$, where

$$\mu^*(A) := \mathbb{P}(Y_1^* \in A) = \mathbb{P}((Y_1^{(1)}, -Y_1^{(2)}) \in A).$$

Define $Z := \left(\sum_{j=1}^{N} Y_j^{(1)}, \sum_{j=1}^{N} Y_j^{(2)}\right) = \sum_{j=1}^{N} (Y_j^{(1)}, Y_j^{(2)})$. Then Z is a compound Poisson random vector with Lévy measure $\nu_Z = \mu - \mu\{0\}\delta_0$, where $\mu(A) = \mathbb{P}((Y_1^{(1)}, Y_1^{(2)}) \in A)$.

$$\begin{aligned}
\nu_{Z^*}(\{x : x_1x_2 > 0\}) &= \mu^*(\{x : x_1x_2 > 0\}) - \mu^*\{0\}\delta_0(\{x : x_1x_2 > 0\}) \\
&= \mu^*(\{x : x_1x_2 > 0\}) \\
&= \mathbb{P}((Y_1^{(1)}, -Y_1^{(2)}) \in \{x : x_1x_2 > 0\}) \\
&= \mathbb{P}(-Y_1^{(1)}Y_1^{(2)} > 0) \\
&= \mathbb{P}(Y_1^{(1)}Y_1^{(2)} < 0) \\
&= \mathbb{P}((Y_1^{(1)}, Y_1^{(2)}) \in \{x : x_1x_2 < 0\}) \\
&= \mu(\{x : x_1x_2 < 0\}) - \mu\{0\}\delta_0(\{x : x_1x_2 < 0\})
\end{aligned}$$

$$= \nu_Z(\{x : x_1 x_2 < 0\})$$
$$> 0$$

where we obtain the inequality from the proof of Theorem 5.4, where ν_Z satisfies (5.23). This gives us our desired result. □

5.3 Negative dependence in Lévy processes

There are situations when condition (5.4) is equivalent to negative association. One case is when X is α-stable, as given in Theorem 5.2. But another case arises in the situation of a Lévy process. First, we define negative association of a stochastic process in the following manner.

Definition 5.6. (a) Let $X = (X_t)_{t\geq 0}$ be a stochastic process. We say X is **negatively associated in space** or **spatially negatively associated** if X_t is negatively associated for all $t \geq 0$.

(b) Let $X = (X_t)_{t\geq 0}$ be a (time-homogeneous) Markov process on $(\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\geq 0}, \mathbb{P}^x)_{x\in\mathbb{R}^d})$. We say X is **negatively associated in space** or **spatially negatively associated** if X_t is negatively associated for all $t \geq 0$ with respect to $\mathbb{P}^x$, for all $x \in \mathbb{R}^d$.

Again we can define NSD, NUOD, NLOD, and NOD in space if we replace "negatively associated" with "NSD," "NUOD," "NLOD," and "NOD," respectively, in the above definitions.

Theorem 5.6. *Let $X = (X_t)_{t\geq 0}$ be a Lévy process in $\mathbb{R}^2$ with characteristics $(b, 0, \nu)$. Then X is spatially negatively associated if and only if ν satisfies (5.4).*

Proof. ($\Leftarrow$). Assume $\nu(\{x : x_1 x_2 > 0\}) = 0$. Then for all $t \geq 0$, $t\nu(\{x : x_1 x_2 > 0\}) = 0$. Since $X_t \sim ID(tb, 0, t\nu)$, by Claim 5.2.1, X_t is NA for all $t \geq 0$.

($\Rightarrow$). Assume X_t is NA for all $t \geq 0$ (wrt $\mathbb{P}^x$ for all x). By Proposition 5.1.3, X_t is also NUOD for all $t \geq 0$ (wrt $\mathbb{P}^x$ for all x). Assume for contradiction that $\nu(\{x : x_1 x_2 > 0\}) = \nu(\mathbb{R}^2_+ \cup \mathbb{R}^2_-) > 0$. WLOG, say $\nu((0,\infty)\times(0,\infty)) > 0$. By continuity of measure and Lemma

3.4.7, there exists $a > 0$ such that $\nu((a,\infty)\times(a,\infty)) > 0$ such that $\nu(\partial((a,\infty)\times(a,\infty))) = 0$ and $\nu(\partial((a,\infty)\times\mathbb{R})) = 0$. We write $\mathbb{P} = \mathbb{P}^0$. Then by (3.27),

$$\begin{aligned}
0 &< \nu((a,\infty)\times(a,\infty)) \\
&= \lim_{t\searrow 0}\frac{1}{t}\mathbb{P}(X_t \in (a,\infty)\times(a,\infty)) \\
&= \lim_{t\searrow 0}\frac{1}{t}\mathbb{P}(X_t^{(1)} > a, X_t^{(2)} > a) \\
&\leq \lim_{t\searrow 0}\frac{1}{t}\mathbb{P}(X_t^{(1)} > a)\mathbb{P}(X_t^{(2)} > a) \\
&= \lim_{t\searrow 0}\frac{1}{t}\mathbb{P}(X_t^{(1)} > a)\cdot\lim_{t\searrow 0}\mathbb{P}(X_t^{(2)} > a) \\
&= \nu((a,\infty)\times\mathbb{R})\cdot\mathbb{P}(X_0^{(2)} > a) \\
&= \nu((a,\infty)\times\mathbb{R})\cdot 0 \\
&= 0
\end{aligned}$$

where we obtain the inequality by NUOD. Thus, we have a contradiction, yielding us our desired result. □

Remark 5.3.1. Once again, we need to revisit this theorem for $d > 2$. Obviously it holds for $d = 2$, but for higher dimensions, the assumption of (5.4) may be too strong. Thus, we can consider projections ν^{ij} as we did in Theorem 5.3.

Theorem 5.7 (Rosinski, Tu, 2017, [40])**.** *Let $X = (X_t)_{t\geq 0}$ be a Lévy process in $\mathbb{R}^d$ with characteristics $(b, 0, \nu)$. Then X is spatially negatively associated if and only if ν satisfies (5.12), i.e. $\nu^{ij}(\mathbb{R}^2_+ \cup \mathbb{R}^2_-) = 0$ for all $i \neq j$.*

Proof. ($\Leftarrow$). Assume condition (5.12). Then by Theorem 5.3, X_1 is NA. Additionally, by Theorem 5.3, $X_t \sim ID(tb, 0, t\nu)$ is NA for all $t \geq 0$.

($\Rightarrow$). Assume X is spatially NA. Then X is also spatially NUOD. Assume for contradiction that there exist $k \neq l$ such that $\nu^{kl}(\mathbb{R}^2_+ \cup \mathbb{R}^2_-) > 0$. WLOG, say $\nu^{kl}((0,\infty)\times(0,\infty)) > 0$. By continuity of measure, there exists $a > 0$ such that $\nu^{kl}((a,\infty)\times(a,\infty)) > 0$

and $\nu^{kl}(\partial[(a,\infty)\times(a,\infty)])=0$ and $\nu^{kl}(\partial[(a,\infty)\times\mathbb{R}])=0$. Define

$$A := \mathbb{R}\times ...\mathbb{R}\times(a,\infty)\times\mathbb{R}\times ...\times\mathbb{R}\times(a,\infty)\times\mathbb{R}\times ...\times\mathbb{R},$$

where "(a,∞)" are located in the kth and l^{th} coordinates, and

$$A' := \mathbb{R}\times ...\times\mathbb{R}\times(a,\infty)\times\mathbb{R}\times ...\times\mathbb{R}\times ...\times\mathbb{R},$$

where "(a,∞)" is located in the kth coordinate. Observe that $\nu(\partial A)=\nu^{kl}(\partial[(a,\infty)\times(a,\infty)])=0$ and $\nu(\partial A')=\nu^{kl}(\partial[(a,\infty)\times\mathbb{R}])=0$. Then by (3.27), we have

$$\begin{aligned}\lim_{t\to0}\frac{1}{t}\mathbb{P}(X_t^{(k)}>a,X_t^{(l)}>a) &= \lim_{t\to0}\frac{1}{t}\mathbb{P}(X_t\in A)\\ &= \nu(A)\\ &= \nu^{kl}((a,\infty)\times(a,\infty))\end{aligned} \tag{5.25}$$

and, similarly,

$$\begin{aligned}\lim_{t\to0}\frac{1}{t}\mathbb{P}(X_t^{(k)}>a,X_t^{(l)}\in\mathbb{R}) &= \lim_{t\to0}\frac{1}{t}\mathbb{P}(X_t\in A')\\ &= \nu(A')\\ &= \nu^{kl}((a,\infty)\times\mathbb{R})\end{aligned} \tag{5.26}$$

By NUOD of X_t, we also have NUOD of $(X_t^{(k)},X_t^{(l)})$ for all $t\geq0$. Then by (5.25), (5.26), and NUOD of $(X_t^{(k)},X_t^{(l)})$ we have

$$\begin{aligned}0 &< \nu^{kl}((a,\infty)\times(a,\infty))\\ &= \lim_{t\to0}\frac{1}{t}\mathbb{P}(X_t^{(k)}>a,X_t^{(l)}>a)\\ &\leq \lim_{t\to0}\frac{1}{t}\mathbb{P}(X_t^{(k)}>a)\mathbb{P}(X_t^{(l)}>a)\\ &= \lim_{t\to0}\frac{1}{t}\mathbb{P}(X_t^{(k)}>a)\lim_{t\to0}\mathbb{P}(X_t^{(l)}>a)\\ &= \nu^{kl}((a,\infty)\times\mathbb{R})\cdot0\\ &= 0,\end{aligned}$$

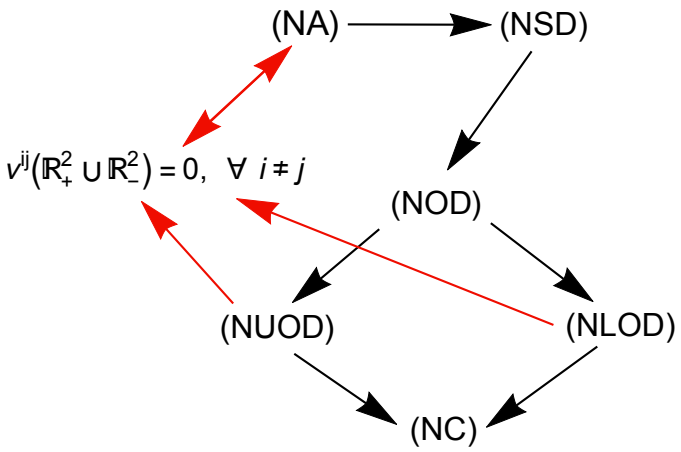

Figure 5.2: Negative dependence for jump-Lévy processes

which yields a contradiction. □

Remark 5.3.2.

(a) The proof of the necessary condition in Theorem 5.7 also gives us that NUOD is equivalent to condition (5.12). Similarly, we can show NLOD is equivalent to (5.12) using the same technique in the above proof. The equivalences can be outlined in the Figure 5.2.

(b) Theorem 5.7 also shows us that pairwise NUOD (or NLOD) in space, i.e. $(X_t^{(i)}, X_t^{(j)})$ is NUOD (or NLOD) for all $i \neq j$, is equivalent to spatial negative association of the Lévy process $(X_t)_{t\geq 0}$.

(c) As we mentioned at the beginning of Section 5.2, in our paper [40], we will establish a much nicer method to characterize positive and negative association (and additional weaker dependence structures) under a unified framework for Lévy processes using small-time asymptotics.

5.4 Positive and negative dependence in limit theorems

Thus far, we've discussed the characterizations of dependence for infinitely divisible distributions and their extensions to Lévy-type Markov processes. However, positively and negatively dependent sequences and processes have interesting applications to limit theorems, particularly "central limit theorems" and "law of large numbers." For example, if a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is associated (or negatively associated), i.e. $(X_1, ..., X_m)$ is an associated (negatively associated) random vector for all $m \in \mathbb{N}$, then under certain conditions, one can say something analogous to the law of large numbers:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) = \mathbb{E}f(X_1), \;\; a.s.$$

and something analogous to the central limit theorem:

$$\frac{S_n - \mu_n}{\sigma_n} \Rightarrow Z,$$

where $S_n = X_1 + ... + X_n$, $(\mu_n)_{n\in\mathbb{N}}$ and $(\sigma_n)_{n\in\mathbb{N}}$ is some re-scaling, and Z is some random variable. In this section, we just give an overview of some of the classical results of dependence in limit theorems, particularly those of Lebowitz [29] and Newman [35], [36], [37]. We also present some recent work in this area.

One of the oldest results of dependence in limit theorems comes from Lebowitz and is an analogous result to the classical "Law of Large Numbers" for an iid sequence. We say a sequence of random variables $X = (X_n)_{n\in\mathbb{N}}$ is stationary if, for all $j \leq n$,

$$(X_j, X_{j+1}, ..., X_n) \stackrel{d}{=} (X_{j+k}, X_{j+1+k}..., X_{n+k}) \tag{5.27}$$

for all $k \in \mathbb{N}$.

Theorem 5.8 (Lebowitz [29]). *Let $X = (X_n)_{n\in\mathbb{N}}$ be a stationary sequence which is associated. If*

$$\lim_{n\to\infty} \frac{\sum_{j=1}^{n} \mathrm{Cov}(X_1, X_j)}{n} = 0, \tag{5.28}$$

then for all $f \in L^1(\mu)$, where $X_1 \sim \mu$, we have

$$\lim_{n\to 0} \frac{\sum_{i=1}^{n} f(X_j)}{n} = \mathbb{E}f(X_1), \ \ a.s. \tag{5.29}$$

If X is instead negatively associated, then (5.28) always holds, and we obtain (5.29).

The classical result of central limit-type theorems for dependent sequences was given by Newman [35], [36]. A crucial result that gives us a central limit theorem for dependent sequences is the following lemma by Newman.

Lemma 5.4.1 (Newman [35]). *Let $X = (X_1, ..., X_d)$ be an associated (or NA) random vector. Then for $u = (u_1, ..., u_d)$,*

$$\left| \mathbb{E}\exp(iu \cdot X) - \prod_{j=1}^{d} \mathbb{E}\exp(iu_j X_j) \right| \leq \sum_{\substack{j,k=1 \\ j<k}}^{d} |u_j u_k \mathrm{Cov}(X_j, X_k)| \tag{5.30}$$

Remark 5.4.1. Newman actually proved Lemma 5.4.1 under weaker dependence notions, i.e. X is *strongly positive orthant dependent (SPOD)* and *strongly negative orthant dependent (SNOD)*, a slightly stronger form of POD and NOD. See Newman [35] or [37] for more details.

The intuitive idea behind the usefulness of this lemma is that we can "approximate" the distribution of a dependent (associated or NA) random vector X by an independent random vector $X^{\perp}$ with the same marginal distributions as X. Thus, since we know about central limit theorems of independent sequences, then we can use the bound in (5.30) to obtain a central limit theorem for the dependent sequence, if we can control the dependence (covariances) of the random variables.

Theorem 5.9 (Newman [35]). *Let $X = (X_n)_{n\in\mathbb{N}}$ be a stationary sequence of random variables which is either associated or negatively associated. Then if*

$$\sigma^2 := Var(X_1) + 2\sum_{j=2}^{\infty} \mathrm{Cov}(X_1, X_j) < \infty \tag{5.31}$$

and $\sigma^2 \geq Var(X_1)$ (if X associated) or $\sigma^2 \in [0, Var(X_1)]$ (if X NA), then

$$\frac{\sum_{i=1}^{n}(X_j - \mathbb{E}X_j)}{\sqrt{n}} \Rightarrow \sigma Z \tag{5.32}$$

where $Z \sim N(0,1)$.

Remark 5.4.2. (a) If X is associated, then assuming $\sum_{j=2}^{\infty} \mathrm{Cov}(X_1, X_j) < \infty$ is sufficient for (5.31). If X is NA, then (5.31) is automatic.

(b) Newman proved Theorem 5.9 under a weaker dependence notion: linearly positive quadrant dependent (LPQD) and linearly negatively quadrant dependent (LNQD). See [35], [37].

Since Newman's papers in 1980s, there have been many improvements and extensions of his results. We highlight one nice result from Cagin et. al. [10], which makes a connection to moderate/large deviations.

Consider an iid sequence $(X_n)_{n\in\mathbb{N}}$ with common law F having mean 0. Let $\sigma_n^2 = \mathbb{E}(S_n^2)$, where $S_n = X_1 + ... + X_n$. The classical central limit theorem, given by $S_n/\sigma_n \Rightarrow Z \sim N(0,1)$, also yields the following approximation of the tails of the sum:

$$\mathbb{P}(S_n > x\sigma_n) \approx \mathbb{P}(Z > x), \;\; n \text{ large} \tag{5.33}$$

for all $x > 0$ fixed. When we let $x = (x_n)_{n\in\mathbb{N}}$ be a sequence such that $x_n \to \infty$, then the approximation of $\mathbb{P}(S_n > x_n s_n)$ becomes a moderate/large deviation problem. We say the problem is a moderate deviation problem if $x_n = O(\sigma_n)$ as $n \to \infty$. Several results have been presented on this approximation problem in the case of an independent sequence $(X_n)_{n\in\mathbb{N}}$. Cagin et. al. [10] presents an approximation for an associated sequence.

For a sequence $(X_n)_{n\in\mathbb{N}}$ we decompose the partial sum $S_n = X_1 + ... + X_n$ into smaller blocks by the following prescription: let $(a_n)_{n\in\mathbb{N}} \subset \mathbb{Z}$ be an increasing sequence such that $a_n < n/2$ for all n. Define $b_n = \max(k \in \mathbb{Z} : k \leq n/2a_n)$, and $Y_{k,n} := \sum_{j=(k-1)a_n+1}^{kp_n} X_j$, for $j = 1, ..., 2b_n$. Thus, we can decompose S_n by

$$S_n = Y_{1,n} + ... + Y_{2b_n,n} + \sum_{j=2b_n a_n+1}^{n} X_j.$$

Define the odd partial sum $W_{n,odd} = \sum_{j=1}^{b_n} Y_{2j-1,n}$.

Theorem 5.10 (Cagin et. al. [10])**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of stationary, mean 0, associated random variables. Assume*

(a) $\mathbb{E}|X_n|^p < \infty$ *for all* n*, for some* $p > 2$*.*

(b) $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}_+$ *such that* $\limsup_{n\to\infty} \frac{x_n}{\log n} = 2\gamma < p - 2$

(c) $\sigma_n^2/n \to \sigma^2$ *for some* $\sigma^2 \in \mathbb{R}_+$*.*

(d) $\sum_{j=n}^{\infty} \mathrm{Cov}(X_1, X_j) = O(n^{-\lambda})$*, where* $\lambda > (1 - 3\gamma)\max(2 + \frac{4}{p-2}, 2 + \frac{4\gamma+2}{p-2\gamma-2})$

(e) $|\mathbb{P}(S_n > x_n s_n) - \mathbb{P}(W_{n,odd} > x_n s_n/\sqrt{2})| = O(n^{-\gamma}(\log n)^{-1/2})$

Then $\mathbb{P}(S_n > x_n s_n) = \mathbb{P}(Z > x_n)(1 + o(1))$*, where* $Z \sim N(0, 1)$*.*

There are many different extensions and applications related to CLT-type theorems and large/moderate deviations. For a nice collection of the classical and recent results in this area see the introduction to [10].

Bibliography

[1] Alam, K. and Saxena, K. (1981). Positive dependence in multivariate distributions. *Communication in Statistics: Theory and Methods*, 10(12):1183–1196. 126

[2] Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge University Press. 37, 38, 40, 43, 46, 47, 48, 49, 50, 52, 63, 69, 90

[3] Applebaum, D. (2007). On the infinitesimal generators of Ornstein-Uhlenbeck processes with jumps in hilbert space. *Potential Anal.*, 26:79–100. 87

[4] Bauerle, N. Blatter, A. and Muller, A. (2008). Dependence properties and comparison results for Lévy processes. *Math. Meth. Oper. Res.*, 67:161–186. 3, 28, 57

[5] Bergmann, R. (1978). Some classes of semi-ordering relations for random vectors and their use for comparing covariances. *Math. Nachr.*, 82:103–114. 25, 26

[6] Böttcher, B. (2010). Feller processes: the next generation in modeling. Brownian motion, Lévy processes and beyond. *PLoS ONE*, 5(12). 3, 46

[7] Böttcher, B. (2013). Feller evolution systems: Generators and approximation. arXiv:1305.0375. 100, 101, 102, 103, 106

[8] Böttcher, B. Schilling, R. and Wang, J. (2013). *Lévy Matters III*. Springer. 46, 51, 55, 91

[9] Bulinski, A. and Shashkin, A. (2007). *Limit Theorems for Associated Random Fields and Related Systems*. Advanced Series on Statistical Science and Applied Probability. 22, 132

[10] Cagin, T. Oliveira, P. and Torrado, N. (2016). A moderate deviation for associated random variables. *Journal of the Korean Statistical Society*, 45(2):285–294. 148, 149

[11] Chen, M. and Wang, F. (1993). On order-preservation and positive correlations for multidimensional diffusion processes. *Probab. Theory Relat. Fields*, 95:421–428. 3, 57, 61, 94

[12] Christofides, T. and E., V. (2004). A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis*, 88:138–151. 31, 131

[13] Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman and Hall. 98, 116, 119

[14] Courrège, P. (1965). Sur la forme intégro-différentielle des opérateurs de C_k^∞ dans C satisfaisant au principe du maximum. *Sém. Théorie du potentiel*, 2. 5, 50, 99

[15] E.B., D. (1965). *Markov Processes*. Springer Berlin Heidelberg. 73

[16] Esary, J.D. Proschan, F. and D.W., W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics*, 38(5):1466–1474. 10, 11, 16, 18, 19, 22

[17] Esary, J.D. Proschan, F. and D.W., W. (2014). Characterization of positively correlated squared Gaussian processes. *Annals of Probability*, 42(2):559–575. 45

[18] Ethier, S. and Kurtz, T. (1986). *Markov Processes - Characterization and Convergence*. Wiley. 50

[19] Figueroa-López, J. (2008). Small-time asymptotics for Lévy processes. *Statistics and Probability Letters*, 78:3355–3365. 65

[20] Fortuin, C. Kastelyn, P. and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103. 17

[21] Harris, T. (1960). A lower bound for the critical probability in a certain percolation process. *Proc. Camb. Phil. Soc.*, 59:13–20. 17

[22] Harris, T. (1977). A correlation inequality for Markov processes in partially ordered state spaces. *Ann. Probab.*, 5(3):451–454. 60, 68

[23] Herbst, I. and Pitt, L. (1991). Diffusion equation techniques in stochastic monotonicity and positive correlations. *Probab. Th. Rel. Fields*, 87:275–312. 61

[24] Hoh, W. (1998). *Pseudo-differential operators generating Markov processes*. PhD thesis, Universitat Bielefeld. 37, 39, 55, 69

[25] Jacod, J. (2007). Asymptotic properties of power variations of Lévy processes. *ESAIM: Probability and Statistics*, 11:173–196. 65

[26] Jacod, J. and Shiryaev, A. (2003). *Limit Theorems for Stochastic Processes*. Springer-Verlag. 51, 117

[27] Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables, with applications. *Annals of Statistics*, 11(1):286–295. 6, 133

[28] Kuhn, F. and Schilling, R. (2016). On the domain of fractional Laplacians and related generators of Feller processes. arXiv:1610.0819v1. 65, 66

[29] Lebowitz, J. (1972). Bounds on the correlations and analyticity properties of ferromagnetic Ising spin systems. *Comm. Math. Phys.*, 28:313–321. 146, 147

[30] Lee, M-L. Rachev, S. and Samorodnitsky, G. (1990). Association of stable random variables. *The Annals of Probability*, 18:1759–1764. 6, 133

[31] Lehmann, E. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, 37(5):1137–1153. 13, 25, 129

[32] Liggett, T. (1985). *Interacting Particle Systems*. Springer. 4, 57, 59, 64

[33] Lindqvist, B. (1987). Monotone and associated Markov chains, with applications to reliability theory. *Journal of Applied Probability*, 24:679–695. 106, 107, 113

[34] Mueller, A. and Stoyan, D. (2002). *Comparison methods for stochastic models and risks*. Wiley. 10, 23, 24, 26, 28, 33, 35, 36, 57, 125, 130, 132

[35] Newman, C. (1980). Normal fluctuations and the FKG inequalities. *Commun. Math. Phys.*, 74:119–128. 146, 147, 148

[36] Newman, C. (1983). A general central limit theorem for FKG systems. *Commun. Math. Phys.*, 91:75–80. 146, 147

[37] Newman, C. (1984). Asymptotic independence and limit theorems for positive and negatively dependent random variables. *IMS Lecture Notes - Monograph Series*, 5:127–140. 146, 147, 148

[38] Pitt, L. (1982). Positive correlated normal random variables are associated. *Ann. Probab.*, 10:496–499. 2, 3, 39

[39] Resnick, S. (1988). Association and extreme value distributions. *Austral. J. Statist.*, 30A:261–271. 2, 3, 13, 40, 42

[40] Rosinski, J. and Tu, E. (2017). Positive and negative dependence structures for infinitely divisible distributions and processes: a unified approach. *In preparation.* 7, 133, 137, 143, 145

[41] Ruschendorf, L. Schnurr, A. and Wolf, V. (2016). Comparison of time-inhomogeneous Markov processes. *Advances in Applied Probability*, 48(4):1015–1044. 24, 122

[42] Ruschendorf, L. (2008). On a comparison result for Markov processes. *J. Appl. Prob.*, 45:279–286. 4, 29, 65, 121

[43] Samorodnitsky, G. (1995). Association of infinitely divisible random vectors. *Stochastic Processes and their Applications*, 55:45–55. 3, 44, 61, 140

[44] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press. 65, 117

[45] Schilling, R. (1998a). Conservativeness and extensions of Feller semigroups. *Positivity*, 2:239–256. 5, 69, 70, 72

[46] Schilling, R. (1998b). Growth and Holder conditions for the sample paths of Feller processes. *Probab. Theory Relat. Fields*, 112:565–611. 5, 62, 64, 69, 72

[47] Schnurr, A. (2009). *The Symbol of a Markov Semimartingale.* PhD thesis, Technischen Universitat Dresden. 37, 38, 51, 55, 63

[48] Schnurr, A. (2013). Generalization of the Blumenthal-Getoor index to the class of homogeneous diffusions with jumps and some applications. *Bernoulli*, 19(5A):2010–2032. 51

[49] Szekli, R. (1995). *Stochastic Ordering and Dependence in Applied Probability.* Springer-Verlag. 10, 65

[50] Tu, E. (2017a). On the association and other forms of positive dependence for Feller processes. *In preparation.* 4, 5, 67, 68, 84

[51] Tu, E. (2017b). Association and other forms of positive dependence for Feller Evolutions Systems. *In preparation.* 6, 107, 111, 115, 120

[52] Wang, J.-M. (2009). Stochastic comparison and preservation of positive correlations for Lévy-type processes. *Acta Mathematica Sinica*, 25(5):741–758. 4, 62, 92

[53] Wang, J.-M. (2013). Stochastic comparison for Lévy-type processes. *J. Theor. Probab.*, 26:997–1019. 93

Vita

Eddie Tu was born in Richmond, VA to parents, Albert and Amy. He has one older sister, Kelly, who is a Professor in Human Development and Family Studies at the University of Illinois, Urbana-Champaign. He attended Hermitage High School and, after graduation, attended Randolph-Macon College in Ashland, VA, where he obtained a B.S. in Mathematics with a minor in English Literature. After graduating from Randolph-Macon in 2011, Eddie attended the University of Tennessee, Knoxville to pursue his graduate studies in Mathematics and work as a graduate teaching assistant. In 2014, he obtained his M.S. in Mathematics, and then began his research in the area of probability and stochastic processes under the guidance of Dr. Jan Rosinski. Eddie will graduate in the summer of 2017 with his Ph.D in Mathematics, along with a M.S. in Statistics. After graduation, he will begin a tenure-track position as assistant professor at Dickinson College in Pennsylvania.