Some Contributions to Distribution Theory and Applications

By

Alessandro Selvitella

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2017) McMaster University
(Statistics) Hamilton, Ontario

TITLE: Some Contributions to Distribution Theory and Applications
AUTHOR: Dr. Alessandro Selvitella
SUPERVISOR: Distinguished University Professor Narayanaswamy Balakrishnan
NUMBER OF PAGES: 286

Abstract

In this thesis, we present some new results in distribution theory for both discrete and continuous random variables, together with their motivating applications.

We start with some results about the Multivariate Gaussian Distribution and its characterization as a maximizer of the Strichartz Estimates. Then, we present some characterizations of discrete and continuous distributions through ideas coming from optimal transportation. After this, we pass to the Simpson's Paradox and see that it is ubiquitous and it appears in Quantum Mechanics as well. We conclude with a group of results about discrete and continuous distributions invariant under symmetries, in particular invariant under the groups A_1, an elliptical version of $O(n)$ and $\mathbb{T}^n$.

As mentioned, all the results proved in this thesis are motivated by their applications in different research areas. The applications will be thoroughly discussed. We have tried to keep each chapter self-contained and recalled results from other chapters when needed.

The following is a more precise summary of the results discussed in each chapter.

In chapter 1, we discuss a variational characterization of the Multivariate Normal distribution (MVN) as a maximizer of the Strichartz Estimates. Strichartz Estimates appear as a fundamental tool in the proof of wellposedness results for dispersive PDEs. With respect to the characterization of the MVN distribution as a maximizer of the entropy functional, the characterization as a maximizer of the Strichartz Estimate does not require the constraint of fixed variance. In this chapter, we compute the precise optimal constant for the whole range of Strichartz admissible exponents, discuss the connection of this problem to Restriction Theorems in Fourier analysis and give some statistical properties of the family of Gaussian Distributions which maximize the Strichartz estimates, such as Fisher Information, Index of Dispersion and Stochastic Ordering. We conclude this chapter presenting an optimization algorithm to compute numerically the maximizers.

Chapter 2 is devoted to the characterization of distributions by means of techniques from Optimal Transportation and the Monge-Ampère equation. We give emphasis to methods to do statistical inference for distributions that do not possess good regularity, decay or integrability properties. For example, distributions which do not admit a finite expected value, such as the Cauchy distribution. The main tool used here is a modified version of the characteristic function (a particular case of the Fourier Transform). An important motivation to develop these tools come from Big Data analysis and in particular the Consensus Monte Carlo Algorithm.

In chapter 3, we study the *Simpson's Paradox*. The *Simpson's Paradox* is the phenomenon that appears in some datasets, where subgroups with a common trend (say, all negative trend) show the reverse trend when they are aggregated (say, positive trend). Even if this issue has an elementary mathematical explanation, the statistical implications are deep. Basic examples appear in arithmetic, geometry, linear algebra, statistics, game theory, sociology (e.g. gender bias in the graduate school admission process) and so on and so forth. In our new results, we prove the occurrence of the *Simpson's Paradox* in Quantum Mechanics. In particular, we prove that the *Simpson's Paradox* occurs for solutions of the *Quantum Harmonic Oscillator* both in the stationary case and in the non-stationary case. We prove that the phenomenon is not isolated and that it appears (asymptotically) in the context of the *Nonlinear Schrödinger Equation* as well. The likelihood of the *Simpson's Paradox* in Quantum Mechanics and the physical implications are also discussed.

Chapter 4 contains some new results about distributions with symmetries. We first discuss a result on symmetric order statistics. We prove that the symmetry of any of the order statistics is equivalent to the symmetry of the underlying distribution. Then, we characterize elliptical distributions through group invariance and give some properties. Finally, we study geometric probability distributions on the torus with applications to molecular biology. In particular, we introduce a new family of distributions generated through stereographic projection, give several properties of them and compare them with the Von-Mises distribution and its multivariate extensions.

Acknowledgements

Foremost, I would like to thank my thesis advisor, Dr. Narayanaswamy Balakrishnan, whose enthusiasm and knowledge inspired and motivated me at every stage of this thesis; for his patient, expert guidance and useful suggestions, I am truly grateful.

I would also like to thank the Department of Mathematics and Statistics at McMaster University, in particular Dr. Maung Min-Oo and Dr. Román Viveros-Aguilera, who have been part of my Supervisory Committee and whose suggestions have definitely improved the content of this thesis.

I thank all my friends who have made these years so special. In particular, I say thank you to Chris Cappadocia and Alberto Chicco who have always helped me in the good and bad moments.

Infine ringrazio la mia mamma Patrizia e i mie nonni Iolanda e Nunzio a cui voglio tantissimo bene e che mi sono sempre vicini anche se sono in Italia (e non li posso vedere quanto vorrei!) insieme al mio nipotino Leonardo (nato proprio oggi!!), and my lovely girlfriend Linda, who is always close to me and I truly love.

Contents

Papers included in the Thesis

This thesis is based on the following papers:

1. A. SELVITELLA, Remarks on the sharp constant for the Schrödinger Strichartz estimate and applications, *Electron. J. Diff. Equ.* Vol. **2015** No. 270 (2015) 1-19. (Chapter 1)

2. A. SELVITELLA, The maximal Strichartz family of Gaussian distributions: Fisher information, index of dispersion, and stochastic ordering, *Int. J. Differ. Equ.* 2016, Art. ID 2343975, 17. (Chapter 1)

3. A. SELVITELLA, On $\frac{1}{\alpha}$-Characteristic Functions and their properties, *Comm. Stat. Th. Meth.*, no. **4**, (2017), 1941-1958. (Chapter 2)

4. A. SELVITELLA, The Monge-Ampère Equation in Transformation Theory and an Application to $\frac{1}{\alpha}$-Probabilities, *Comm. Stat. Th. Meth.*, no. **4** (2017) 2037-2054. (Chapter 2)

5. A.SELVITELLA, The ubiquity of the Simpson's Paradox, *J. Stat. Distrib. and Appl.* **4:2** (2017). (Chapter 3)

6. A.SELVITELLA, The Simpson's Paradox in Quantum Mechanics, *J. Math. Phys.* **58** no. 3 (2017) 032101 37 pp. (Chapter 3)

7. N. BALAKRISHNAN AND A. SELVITELLA, Symmetry of a Distribution via Symmetry of Order Statistics, *Statistics and Probability Letters* **129** (October 2017) 367-372. (Chapter 4)

8. L. FORSON AND A. SELVITELLA, A simple Characterization of Elliptical Distributions through Group Invariance, submitted. (Chapter 4)

9. A. SELVITELLA, On Geometric Probability Distributions on the Torus with Applications to Molecular Biology, submitted. (Chapter 4)

Chapter 1

Variational Characterization of the Multivariate Normal distribution through Strichartz Norms

The Multivariate Normal (MVN) distribution has several different characterizations. We refer to [47], [90], [51] and [72] for more details about the MVN.

In this chapter, we concentrate on those characterizations of the MVN through variational principles. We start with the well known characterization via entropy maximization under the constraint of fixed variance and pass later to our new results related to the maximization of Strichartz Estimates.

1.1 Characterization through Entropy

A well known characterization of the Gaussian distribution is through the maximization of the *Differential Entropy*, under the constraint of fixed variance Σ. We focus on the case of when the support of the probability density function (pdf) is the whole Euclidean Space $\mathbb{R}^n$.

Theorem 1.1.1. *Let X be a random vector whose pdf is f_X. The* Differential Entropy *$h(X)$ is defined by the following functional:*

$$h(X) := -\int_{\mathbb{R}^n} f_X(x) \log f_X(x)\, dx.$$

The Multivariate Normal Distribution *has the largest* Differential Entropy *$h(X)$ amongst* *all the random variables X with equal variance Σ. Moreover, the maximal value of the* Differential Entropy $h(X)$ *is* $h(MVN(\Sigma)) = \frac{1}{2}log[2\pi e|\Sigma|]$.

The proof is very simple and can be found in several places (see for example a nice treatment in [33]). For completeness, we report the computation in the case $n = 1$.

1.1.1 Proof of Theorem 1.1.1

We consider the variational derivative of $h(X)$ with the constraint of f being a probability distribution and with the constraint of having a fixed variance σ^2. This gives rise to the following *Euler Lagrange Equation* with two *Lagrangian multipliers* λ_0 and λ_2:

$$\mathcal{L}(v;\lambda_0,\lambda_2) = -\int_{\mathbb{R}} v(x)\ln(v(x))\,dx + \lambda_0\left(1 - \int_{\mathbb{R}} v(x)\,dx\right) + \lambda_2\left(\sigma^2 - \int_{\mathbb{R}} v(x)(x-\mu)^2\,dx\right)$$

with $v(x)$ being some function with *Expected Value* $\mu := \int_{\mathbb{R}} xv(x)dx$. The two *Lagrangian multipliers* λ_0 and λ_2 appear, because of the two constraints. One constraint is related to the normalization condition

$$\int_{\mathbb{R}} v(x)dx = 1$$

and the other is related to the requirement of fixed variance

$$\sigma^2 = \int_{\mathbb{R}} v(x)(x-\mu)^2 dx.$$

Now, we take the variational derivative of the functional $\mathcal{L}$. To be at a critical point, we need to impose that this variational derivative is zero for every test function $g(x)$. Therefore, we get:

$$0 = -\mathcal{L}'(v)g = \frac{d}{dt}|_{t=0} L(v(x) + tg(x)) = \int_{-\infty}^{\infty} g(x)\left(\ln(v(x)) + 1 + \lambda_0 + \lambda_2(x-\mu)^2\right)\,dx.$$

Since this must hold for any variation $g(x)$, the term in brackets must be zero, and so, solving for $v(x)$, it yields:

$$v(x) = e^{-\lambda_0 - 1 - \lambda_2(x-\mu)^2}.$$

Now, we use the constraint of the problem and solve for λ_0 and λ_2. From $\int_{\mathbb{R}} v(x)\,dx = 1$, we get the condition

$$\lambda_2^{-\frac{1}{2}}\pi^{\frac{1}{2}} = e^{\lambda_0+1}$$

and from $\sigma^2 = \int_{\mathbb{R}} v(x)(x-\mu)^2 dx$, we get

$$\sigma^2 e^{\lambda_0+1}\lambda_2^{\frac{3}{2}} = \pi^{\frac{1}{2}}/2.$$

Solving for λ_0 and λ_2 we get $\lambda_0 = \frac{1}{2}log(2\pi\sigma^2) - 1$ and $\lambda_2 = \frac{1}{2\sigma^2}$ which altogether give the *Gaussian Distribution*:

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

A similar argument for the second derivative, gives

$$< \mathcal{L}''(v)g_1, g_2 >= -\int_{S\subset\mathbb{R}} g_1(x)v(x)^{-1}g_2(x)dx$$

with S the support of $v(x)$ and $g_1(x), g_2(x)$ test functions. Negativity ($v(x) > 0$ for every $x \in \mathbb{R}$) gives maximality. To get the optimal entropy constant, it is enough to plug inside the optimizer in $h(X)$.

1.2 A Remark on the Sharp Constant for the Schrödinger Strichartz Estimate and Applications

In this section, we compute the sharp constant for the Homogeneous Schrödinger Strichartz Inequality and Fourier Restriction Inequality on the Paraboloid in any dimension under the condition, as it is conjectured (and proved in dimensions $n = 1$ and $n = 2$), that the maximizers are Gaussians.

We observe also how this would imply a far from optimal, but "cheap" and sufficient, criterion of global wellposedness in the L^2-critical case $p = 1 + 4/n$. With respect to the characterization of the MVN through entropy maximization, here, there is no constraint of fixed variance required. The results of this section highlight how the MVN plays a fundamental role in fields very different from probability and statistics.

1.2.1 Introduction and Motivation

Consider the following Nonlinear Schrödinger Equation (from now on NLS):

$$i\partial_t u(t,x) + \Delta u(t,x) + \mu|u|^{p-1}u(t,x) = 0 \quad (t,x) \in (0,\infty)\times\mathbb{R}^n, \tag{1.1}$$

with initial datum $u(0,x) = u_0(x), \quad x \in \mathbb{R}^n$. Here, the space dimension is $n \geq 1$, the nonlinearity has $p \geq 1$ and $\mu = -1, 0, 1$ in which cases the equation is said to be *defocusing*, *linear* and *focusing*, respectively.

Extended research has been done to prove the global wellposedness of the above problem in the scale of Hilbert Spaces $H^s(\mathbb{R}^n)$ (see Section 1.2.2 for a precise definition). In the case of regular solutions $s > n/2$, the *algebra property* of the space $H^s(\mathbb{R}^n)$ makes the proof simpler, while in the case $s \leq n/2$ one needs Strichartz estimates to close the argument (see again Section 1.2.2). We refer to [116] for more details and references.

Strichartz Estimates were originally proved by Strichartz [113] in the non end-point case and much later for the end-point case by Keel and Tao [68] in the homogeneous case and by Foschi [54] in the inhomogeneous case, following Keel and Tao's approach. See also [114] After Strichartz's work, a huge research field opened and Strichartz estimates were proved for several different equations. See [116] and the references therein, for a more complete discussion on Strichartz Estimates.

Several mathematicians have then been interested in the problem of the sharpness of Strichartz Inequalities. As far as we know, the first one addressing this problem has been Kunze [74], who proved the existence of a maximizing function for the estimate $||e^{it\partial_x^2}u||_{L^6_{t,x}(\mathbb{R}^2)} \leq S_h(1)||u||_{L^2}(\mathbb{R})$ (case of dimension $n = 1$), by means of the concentration compactness principle used in the Fourier Space and by means of multilinear estimates due to Bourgain [18]. This method has been first developed by him in relation to a variational problem from nonlinear fiber optics on Strichartz-type Estimates [73]. The first author to give explicit values of the sharp Strichartz Constants and characterize the maximizers has been Foschi [53], who proved that in dimension $n = 1$ the sharp constant is $S_h(1) = 12^{-1/12}$, while in dimension $n = 2$ the sharp constant is $S_h(2) = 2^{-1/2}$. He also proved that the maximizer is the Gaussian function $f(x) = e^{-|x|^2}$ (up to symmetries) in both dimensions $n = 1$ and $n = 2$ (see Section 1.2.2 below). He moreover conjectured (Conjecture 1.10) that Gaussians are maximizers in every dimension $n \geq 1$. Independently, this result has been reached also by Hundertmark and Zharnitsky in [63] that gave also a conjecture on the value of the Strichartz Constant (Conjecture 1.7). An extension of these results can be found in [24]. A step towards proving Foschi's conjecture has been done by Christ and Quilodán [27], who demonstrated that Gaussians are critical points in any dimension $n \geq 1$. They do not give any conjecture on the explicit value of the sharp Strichartz Constant $S_h(n)$ for general dimension n. Duyckaerts, Merle and Roudenko in [46] give an estimate of $S_h(n)$ and also precise asymptotics in the small data regime, but not the explicit value.

Here, assuming that Gaussians are actually maximizers, as it is conjectured, and not just critical points, we compute the Strichartz Constant in a setting a little more general than the one of the conjecture of Hundertmark and Zharnitsky [63] and this is the main contribution of the section.

Theorem 1.2.1. *Suppose Gaussians maximize Strichartz Estimates for any $n \geq 1$. Then, for any $n \geq 1$ and (q,r) admissible pair (see Section 1.2.2 below), the* sharp Homogeneous Strichartz Constant *$S_h(n,q,r) = S_h(n,r)$ defined by*

$$S_h(n,r) := \sup \left\{ \frac{||u||_{L^q_t L^r_x(\mathbb{R}\times\mathbb{R}^n)}}{||u||_{L^2_x(\mathbb{R}^n)}} : u \in L^2_x(\mathbb{R}^n), u \neq 0 \right\}, \tag{1.2}$$

is given by

$$S_h(n,r) = 2^{\frac{n}{4} - \frac{n(r-2)}{2r}} r^{-\frac{n}{2r}}. \tag{1.3}$$

Moreover, if we define $S_h(n) := S_h(n, 2+4/n, 2+4/n)$ by

$$S_h(n) = \sup \left\{ \frac{||u||_{L^{2+4/n}_{t,x}(\mathbb{R}\times\mathbb{R}^n)}}{||u||_{L^2_x(\mathbb{R}^n)}} : u \in L^2_x(\mathbb{R}^n), u \neq 0 \right\}, \tag{1.4}$$

then for every $n \geq 1$ we have that

$$S_h(n) = \left(\frac{1}{2}(1+\frac{2}{n})^{-n/2} \right)^{\frac{1}{2+4/n}}; \tag{1.5}$$

$S_h(n)$ is a decreasing function of n and

$$S_h(n) \to \frac{1}{(2e)^{1/2}}, \quad n \to +\infty.$$

For any $n \geq 1$ and $(\tilde{q},\tilde{r})$ admissible pair, the sharp Dual Homogeneous Strichartz Constant *$S_d(q,r,n) = S_d(n,r)$ is defined by*

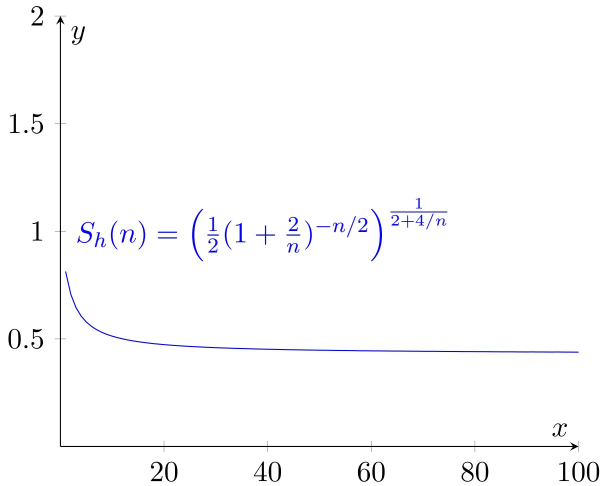
$$S_d(n,r) := \sup \left\{ \frac{\left|\left|\int_{\mathbb{R}} e^{is\Delta} F(s)ds\right|\right|_{L^2_x}}{||F||_{L^{\tilde{q}'}_t L^{\tilde{r}'}_x}} : F \in L^{\tilde{q}'}_t L^{\tilde{r}'}_x(\mathbb{R}\times\mathbb{R}^n), F \neq 0 \right\}, \tag{1.6}$$

We have that $S_h(n,r) = S_d(n,r)$.

Remark 1.2.2. *We notice that q and r are not independent since they are an admissible pair. For this reason, q appears in $S(n,r)$ just as a function of r. One could have also expressed the sharp constant as a function of q by*

$$S_h(n,q) = 2^{-\frac{1}{q}} \left(1 - \frac{4}{qn}\right)^{-1/q+n/4},$$

since $r = \frac{2qn}{nq-4}$ (just plug this expression inside $S_h(n,r)$).



Remark 1.2.3. *We can see that, for $n = 1$ and $n = 2$, we recover the values of $S_h(n)$ found by Foschi in [53].*

Remark 1.2.4. *The asymptotics of $S_h(n)$ basically say that in the non-compact case of $\mathbb{R}^n$, the increase of the spatial dimension n allows more dispersion, but the rate of dispersion, measured by the Homogeneous Strichartz Estimate, does not increase indefinitely. We believe that a similar phenomenon should appear in the case of the Schrödinger equation on the hyperbolic space. We think that it might not be the case for manifolds which become more and more negatively curved with the increase of the dimension, in which case we might observe an indefinitely growing dispersion rate.*

The knowledge of the Optimal Strichartz Constant gives a more precise upper bound on the size of the $L^2 - norm$ for which the "cheapest argument" (the standard Duhamel's Principle -see for example [116]-) gives global wellposedness for (1.1) in the L^2-critical case $p = 1 + 4/n$. From now on we will concentrate on the case $s = 0$ (note $0 < n/2$ for every $n > 0$), namely we will consider just the case in which the initial datum $u_0(x) \in L^2(\mathbb{R}^n)$ and just the case of not *supercritical* nonlinearities $1 < p \leq 1+4/n$. In the *subcritical* case $1 < p < 1+4/n$, Tsustsumi [119] proved local wellposedness and also global wellposedness due to the fact that the local time of existence given by his strategy depends just on the L^2-norm of the initial datum and that the NLS have a conservation law at the L^2-regularity ($T_{loc} = T_{loc}(||u_0||_{L^2(\mathbb{R}^n)})$). Also in the *critical* case, Tsutsumi proved local wellposedness, thanks to the global bound of the $L^{2(n+2)/n}_{t,x}$ Strichartz Norm (see Section 1.2.2), but now the conservation law could not lead to global existence because the local existence time depends on the profile of the solution ($T_{loc} = T_{loc}(u_0)$). The problem of global wellposedness for the NLS, in the L^2-critical case in any dimension, has been solved just recently in a series of papers by Dodson (see [43], [44], [45]). However if the initial datum is "sufficiently small" in L^2_x then one can get global existence with the argument developed in [119], namely by a straight contraction mapping argument. Here, we give a more precise estimate of this "sufficiently small" and so we have the following theorem.

Theorem 1.2.5. *Consider equation (1.1) with initial datum $u_0(x) \in L^2_x(\mathbb{R}^n)$ satisfying the following bound*

$$||u_0(x)||_{L^2_x} < \frac{1}{S_h(n,r)\alpha}\left(\frac{1}{S_i(n,r)} - \frac{1}{S_i(n,r)\alpha}\right)^{n/4} \tag{1.7}$$

with $\alpha = 2$ if $n \geq 4$ and $\alpha = 1 + n/4$ for $1 \leq n \leq 4$. Here $S_h(n,r)$ and $S_i(n,r)$ are, respectively, the sharp Homogeneous and Inhomogeneous Strichartz Constants. Then, there is a unique global solution $u(t,x) \in L^2_x(\mathbb{R}^n)$ for every $t \geq 0$.

Remark 1.2.6. *This result reminds a bit what happens in the focusing case, in which there is an upper bound on the size of the L^2-norm of the initial datum for which one can get global well-posedness and condition (1.7) looks like the Gagliardo-Nirenberg Inequality (see [122] and [116]). Anyways, we want to make clear that condition (1.7) is in some sense fictitious and it is not a threshold. See, for example, the results of Dodson [43], [44], [45].*

Strichartz Inequalities can be set in the more general framework of Fourier Restriction Inequalities in Harmonic Analysis. This connection has been made already clear in the original paper of Strichartz [113]. Therefore, Theorem 1.2.1 can be rephrased in this framework.

Theorem 1.2.7. *Fix $n \geq 1$ and consider the paraboloid $(\mathbf{P}^n, dP^n)$ defined in (1.19) and (1.20) below. Suppose Gaussians maximize the Fourier Restriction Inequality*

$$||\widehat{fdP^n}||_{L_{t,x}^{\frac{2(n+2)}{n}}(\mathbb{R}^{n+1})} \leq S_h(n)||f||_{L^2(\mathbf{P}^n,dP^n)} \tag{1.8}$$

Then, the sharp constant $S_h(n)$ is given by

$$S_h(n) = \Big(\frac{1}{2}(1+\frac{2}{n})^{-n/2}\Big)^{\frac{1}{2+4/n}}.$$

The remaining part of the section is organized as follows. In Subsection 1.2.2, we fix some notation and collect some preliminary results, about the Fourier Transform and the Fundamental Solution for the Linear Schrödinger Equation, about the Strichartz Estimates and their symmetries and the main results in the literature about maximizers for the Strichartz Inequality and about the sharp Strichartz Constant. In Subsection 1.2.3, we prove Theorem 1.2.1, while, in Section 1.2.4, we prove Theorem 1.2.5. In Subsubsection 1.2.5, we discuss the connection between Strichartz and Restriction Inequalities, proving Theorem 1.2.7 in Subsubsection 1.2.5.1. In the Appendix, we give some further comments on the Inhomogeneous Strichartz estimate and on the wave equation.

1.2.2 Notations and Preliminaries

With Schwartz functions we will mean functions belonging to the following function space

$$\mathcal{S}\left(\mathbb{R}^n\right) := \left\{ f \in C^\infty(\mathbb{R}^n) \mid \|f\|_{\alpha,\beta} < \infty \quad \forall \alpha, \beta \right\},$$

with α and β multi-indices, endowed with the following norm

$$\|f\|_{\alpha,\beta} = \sup_{x\in\mathbb{R}^n} \left| x^\alpha D^\beta f(x) \right|.$$

Let (X, Σ, μ) be a measure space. For $1 \le p \le +\infty$, we define the space $L^p(X)$ of all measurable functions from $f : X \to \mathbb{C}$ such that

$$||f||_{L^p(X)} := \left(\int_X |f|^p \, d\mu \right)^{1/p} < \infty.$$

Consider $f : \mathbb{R}^n \to \mathbb{C}$ a Schwartz function in space and $F(t,x) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{C}$ a Schwartz function in space and time. We well use the following notation (and constants) for the space Fourier Transform

$$\hat{f}(\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x) dx$$

and for the Inverse space Fourier Transform

$$f(x) := \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \hat{f}(\xi) d\xi,$$

and the following for the space-time Fourier Transform

$$\mathcal{F}(F)(\tau,\xi) := \frac{1}{(2\pi)^{\frac{n+1}{2}}}\int_{\mathbb{R}^n} e^{-it\tau-ix\cdot\xi}f(t,x)dxdt$$

and the Inverse space-time Fourier Transform

$$F(t,x) := \frac{1}{(2\pi)^{\frac{n+1}{2}}}\int_{\mathbb{R}^{n+1}} e^{it\tau+ix\cdot\xi}\mathcal{F}(\tau,\xi)d\xi d\tau.$$

By means of the Fourier Transform, we can finally define H^s-spaces as the set of functions f such that

$$||f||_{H^s(\mathbb{R}^n)} := \left(\int_{\mathbb{R}^n}|\hat{f}(\xi)|^2(1+|\xi|)^{2s}\right)^{\frac{1}{2}} < +\infty,$$

with $s \in \mathbb{R}$. Mixed spaces such as $L^q_tL^r_x(\mathbb{R}\times\mathbb{R}^n)$ include functions f such that

$$\|f\|_{L^q_tL^r_x(\mathbb{R}\times\mathbb{R}^n)} := \|\|f\|_{L^r_x(\mathbb{R}^n)}\|_{L^q_t(\mathbb{R})} < +\infty$$

with $q, t \in \mathbb{R}$.

1.2.2.1 The Fourier Transform and the Fundamental Solution for The Linear Schrödinger Equation

In this subsection we solve the Linear Schrödinger Equation

$$i\partial_t u(t,x) = \Delta u(t,x), \quad (t,x) \in (0,\infty)\times\mathbb{R}^n, \tag{1.9}$$

with initial datum $u_0(x) = e^{-|x|^2} \in \mathcal{S}(\mathbb{R}^n)$. These computations are well known, but we will rewrite them here in order to clarify what we will compute in the next sections. Since $u_0(x) \in \mathcal{S}(\mathbb{R}^n)$, then also $\partial_t u(t,x) \in \mathcal{S}(\mathbb{R}^n)$ and $\Delta u(t,x) \in \mathcal{S}(\mathbb{R}^n)$. So we can apply the Fourier Transform to both sides of (1.9) and get:

$$i\hat{u}_t = -|\xi|^2\hat{u},$$

whose solution is

$$\hat{u}(\xi,t) = e^{i|\xi|^2t}\hat{u}(\xi,0).$$

So we just need to compute the Fourier Transform of the initial datum and then the Inverse Fourier Transform of $\hat{u}(t,\xi)$ to get the explicit form of the solution.

$$\begin{aligned}\hat{u}(0,\xi) \quad &= \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n} e^{-ix\cdot\xi}u(0,x)dx = \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n} e^{-ix\cdot\xi}e^{-|x|^2}dx \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n} e^{-(|x|^2+ix\cdot\xi-|\xi|^2/4)}e^{-|\xi|^2/4}dx = \frac{e^{-|\xi|^2/4}}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n} e^{-|x-i\xi/2|^2}dx.\end{aligned}$$

by using contour integrals. We notice that, with a simple change of variables, we have:

$$2^{n/2}\int_{\mathbb{R}^n} e^{-|x-i\xi/2|^2}dx = 2^{n/2}\int_{\mathbb{R}^n} e^{-|x|^2}dx = \int_{\mathbb{R}^n} e^{-|x|^2/2}dx = (2\pi)^{n/2}.$$

Hence

$$\hat{u}(0,\xi) = \frac{e^{-|\xi|^2/4}}{(2\pi)^{\frac{n}{2}}}\pi^{n/2} = \frac{e^{-|\xi|^2/4}}{2^{\frac{n}{2}}}.$$

With this we can conclude:

$$\begin{aligned}u(t,x) \quad &= \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n} e^{i|\xi|^2t+ix\cdot\xi}\frac{e^{-|\xi|^2/4}}{2^{\frac{n}{2}}} = \frac{1}{2^n}\frac{1}{\pi^{n/2}}\int_{\mathbb{R}^n} e^{-|\xi|^2(1/4-it)+ix\dot{\xi}}d\xi \\ &= \frac{1}{2^n}\frac{1}{\pi^{n/2}}\int_{\mathbb{R}^n} e^{-(|\xi|^2(1/4-it)-ix\dot{\xi}-|x|^2/(1-4it))}e^{-|x|^2/(1-4it)}d\xi = \\ &= \frac{1}{2^n}\frac{1}{\pi^{n/2}}e^{-|x|^2/(1-4it)}\int_{\mathbb{R}^n} e^{-\left|\xi\sqrt{1/4-it}+ix/(\sqrt{1-4it})\right|^2}d\xi.\end{aligned}$$

Now we make the change of variables $\eta = \xi\sqrt{1/4-it}+ix/(\sqrt{1-4it})$ to get

$$\begin{aligned}u(t,x) \quad &= \frac{1}{2^n}\frac{1}{\pi^{n/2}}e^{-|x|^2/(1-4it)}\int_{\mathbb{R}^n} e^{-|\eta|^2}(1/4-it)^{-n/2}d\eta \\ &= \frac{1}{2^n}\frac{1}{\pi^{n/2}}e^{-|x|^2/(1-4it)}(1/4-it)^{-n/2}\pi^{n/2} = (1-4it)^{-n/2}e^{-\frac{|x|^2}{1-4it}}\end{aligned}$$

Hence

$$u(t,x) = (1-4it)^{-n/2}e^{-\frac{|x|^2}{1-4it}}. \tag{1.10}$$

1.2.2.2 Strichartz Estimates and their symmetries

In this subsection, we state the Strichartz Estimates for the Schrödinger equation, since they are the main topic of the present section and it will help to clarify the statement of our main theorems.

Definition 1.2.8. *Fix* $n \geq 1$. *We call a set of exponents* (q, r) admissible *if* $2 \leq q, r \leq +\infty$ *and*

$$\frac{2}{q} + \frac{n}{r} = \frac{n}{2}.$$

Proposition 1.2.9. *[113], [68], [54] Suppose* $n \geq 1$. *Then, for every* (q, r) *and* $(\tilde{q}, \tilde{r})$ *admissible and for every* $u_0 \in L^2_x(\mathbb{R}^n)$ *and* $F \in L^{\tilde{q}'}_t L^{\tilde{r}'}_x(\mathbb{R} \times \mathbb{R}^n)$, *the following hold:*

- *the* Homogeneous Strichartz Estimates

$$||e^{-it\Delta}u_0||_{L^q_t L^r_x} \leq S_h(n, q, r)||u_0||_{L^2_x};$$

- *the* Dual Homogeneous Strichartz Estimates

$$\left|\left|\int_{\mathbb{R}} e^{is\Delta}F(s)ds\right|\right|_{L^2_x} \leq S_d(n, q, r)||F||_{L^{\tilde{q}'}_t L^{\tilde{r}'}_x};$$

- *the* Inhomogeneous Strichartz Estimates

$$\left|\left|\int_{s<t} e^{-i(t-s)\Delta}F(s)ds\right|\right|_{L^q_t L^r_x} \leq S_i(n, q, r, \tilde{q}, \tilde{r})||F||_{L^{\tilde{q}'}_t L^{\tilde{r}'}_x}.$$

As explained for example in [53], Strichartz Estimates are invariant by the following set of symmetries.

Lemma 1.2.10. *[53] Let* $\mathcal{G}$ *be the group of transformations generated by:*

- *space-time translations:* $u(t, x) \mapsto u(t + t_0, x + x_0)$, *with* $t_0 \in \mathbb{R}$, $x_0 \in \mathbb{R}^n$;
- *parabolic dilations:* $u(t, x) \mapsto u(\lambda^2 t, \lambda x)$, *with* $\lambda > 0$;
- *change of scale:* $u(t, x) \mapsto \mu u(t, x)$, *with* $\mu > 0$;
- *space rotations:* $u(t, x) \mapsto u(t, Rx)$, *with* $R \in SO(n)$;
- *phase shifts:* $u(t, x) \mapsto e^{i\theta}u(t, x)$, *with* $\theta \in \mathbb{R}$;
- *Galilean transformations:*

$$u(t, x) \mapsto e^{\frac{i}{4}\left(|v|^2 t + 2v\cdot x\right)}u(t, x + tv),$$

  with $v \in \mathbb{R}^n$.

Then, if u solves equation (1.9) and $g \in \mathcal{G}$, also $w = g \circ u$ solves equation (1.9). Moreover, the constants $S_h(n,q,r)$, $S_d(n,q,r)$ and $S_i(n,q,r,\tilde{q},\tilde{r})$ are left unchanged by the action of $\mathcal{G}$.

Remark 1.2.11. *For Strichartz Estimates for different equations and different regularities, we refer to [116].*

1.2.2.3 Previous Results on Sharp Strichartz Constant and Maximizers

Here, we collect the results concerning the optimization of Strichartz Inequalities that we need for the next sections. For a broader discussion, we refer to [117] and the references therein.

Proposition 1.2.12. *[74], [27], [53] For any $n \geq 1$ and (q,r) admissible pair, we define $S_h(n) := S_h(n, 2+4/n, 2+4/n)$ by*

$$S_h(n) := \sup \left\{ \frac{||u||_{L^{2+4/n}_{t,x}(\mathbb{R}\times\mathbb{R}^n)}}{||u||_{L^2_x(\mathbb{R}^n)}} : u \in L^2_x(\mathbb{R}^n), u \neq 0 \right\}. \tag{1.11}$$

Then we have the following results:

- *Radial Gaussians are critical points of the Homogeneous Strichartz Inequality in any dimension $n \geq 1$ for all admissible pairs $(q,r) \in (0,+\infty) \times (0,+\infty)$;*
- *The sharp Strichartz Constants $S_h(n)$ can be computed explicitly in dimension $n = 1$: $S_h(1) = 12^{-1/12}$; and dimension $n = 2$: $S_h(2) = 2^{-1/2}$. Moreover, in both the cases $n = 1$ and $n = 2$, the maximizers are Gaussians.*

1.2.3 Proof of Theorem 1.2.1

We are ready to prove Theorem 1.2.1. We assume, as conjectured, that radial Gaussians are mazimizers and not just critical points as proved in [27]. So we will take $u_0(x) = e^{-|x|^2}$. By Lemma 1.2.10, the choice of the Gaussian is done without loss of generality. We start to compute the L^2-norm of the initial datum and so of the solution:

$$||u(t,x)||_{L^2_x} = ||u_0(x)||_{L^2_x} = \Big(\int_{\mathbb{R}^n} e^{-2|x|^2} dx \Big)^{1/2} = \Big(\int_{\mathbb{R}^n} e^{-2|x|^2/4} 2^{-n} dy \Big)^{1/2} \tag{1.12}$$

$$= 2^{-n/2} \Big(\int_{\mathbb{R}^n} e^{-|x|^2/2} dy \Big)^{1/2} = 2^{-n/2} (2\pi)^{n/4} = \Big(\frac{\pi}{2}\Big)^{\frac{n}{4}} \tag{1.13}$$

by similar computations as in Subsection 1.2.2.1.
Now we compute the $L_t^q L_x^r$-norm of the linear solution

$$u(t,x) = (1-4it)^{n/2} e^{-\frac{|x|^2}{1-4it}}.$$

First

$$|u(t,x)|^r = |1-4it|^{-rn/2} \left| e^{-\frac{|x|^2}{1-4it}} \right|^r = |1+16t^2|^{-rn/4} \left| e^{-\frac{(1+4it)|x|^2}{1+16t^2}} \right|^r = |1+16t^2|^{-rn/4} e^{-\frac{r|x|^2}{1+16t^2}}.$$

Then

$$||u(t,x)||_{L_x^r}^r = |1+16t^2|^{-rn/4} \int_{\mathbb{R}^n} e^{-\frac{r|x|^2}{1+16t^2}} dx$$

By the change of variable

$$y = r^{1/2}(1+16t^2)^{-1/2}$$

and hence $dy = r^{n/2} x (1+16t^2)^{-n/2} dx$, we get

$$||u(t,x)||_{L_x^r}^r = |1+16t^2|^{n/2-rn/4} r^{-n/2} \int_{\mathbb{R}^n} e^{-|y|^2} dy = |1+16t^2|^{n/2-rn/4} r^{-n/2} \pi^{n/2},$$

which implies

$$||u(t,x)||_{L_x^r} = |1+16t^2|^{n/(2r)-n/4} r^{-n/(2r)} \pi^{n/(2r)}.$$

Now we have to take the L_t^q-norm of what we obtained:

$$||u(t,x)||_{L_t^q L_x^r} = \Big(\int_{\mathbb{R}^n} ||u(t,x)||_{L_x^r}^q \Big)^{1/q}$$

which means, since (q,r) is an admissible pair (and so $q = 4r/[n(r-2)]$), that

$$||u(t,x)||_{L_t^q L_x^r} = \Big(\int_{\mathbb{R}^n} ||u(t,x)||_{L_x^r}^{\frac{4r}{n(r-2)}} \Big)^{\frac{n(r-2)}{4r}} = \Big[\int_{\mathbb{R}} |1+16t^2|^{-1} \Big]^{\frac{n(r-2)}{4r}} \left(\frac{\pi}{r}\right)^{n/(2r)},$$

since $(n/(2r)-n/4)q = -1$. Now by a simple change of variable inside the integral $(4t = s)$ we get:

$$||u(t,x)||_{L_t^q L_x^r} = \left(\frac{\pi}{r}\right)^{\frac{n}{2r}} \left(\frac{\pi}{4}\right)^{\frac{n(r-2)}{4r}}.$$

Putting everything together we get the equation:

$$S(n,r)\left(\frac{\pi}{2}\right)^{n/4} = \left(\frac{\pi}{r}\right)^{\frac{n}{2r}}\left(\frac{\pi}{4}\right)^{\frac{n(r-2)}{4r}}$$

and so

$$S(n,r) = 2^{\frac{n}{4}-\frac{n(r-2)}{2r}} r^{-\frac{n}{2r}}.$$

In the case $q = r = 2 + 4/n$ one gets:

$$||u(t,x)||^q_{L^q_{t,x}} = q^{-n/2}\pi^{n/2}\int_{\mathbb{R}} |1+16t^2|^{-1} = \pi^{n/2}(2+4/n)^{-n/2}\frac{\pi}{4}.$$

Putting all the information together we get:

$$2^{-2}\pi^{1+n/2}(2+4/n)^{-n/2} = S_h(n)^{2+4/n}(\pi/2)^{1+n/2}$$

and solving for $S_h(n)$ one gets:

$$S(n) = \left(\frac{1}{2}\left(1+\frac{2}{n}\right)^{-n/2}\right)^{\frac{1}{2+4/n}}$$

Now we have to prove that $S_h(n)$ is a decreasing function of n, namely we have to prove that:

$$\left(\frac{1}{2}\left(1+\frac{2}{n+1}\right)^{-(n+1)/2}\right)^{\frac{1}{2+4/(n+1)}} = S_h(n+1) \leq S_h(n) = \left(\frac{1}{2}\left(1+\frac{2}{n}\right)^{-n/2}\right)^{\frac{1}{2+4/n}}.$$

Taking the natural logarithm to both sides and using the fact that the logarithm is a monotone increasing function of his argument we get:

$$\frac{1}{2+4/(n+1)}\Big[-log(2)-\frac{n+1}{2}log(1+2/(n+1))\Big] \leq \frac{1}{2+4/n}\Big[-log(2)-\frac{n}{2}log(1+2/n)\Big].$$

We can easily see that

$$\frac{-log(2)}{2+4/(n+1)} \leq \frac{-log(2)}{2+4/n},$$

so it remains to prove that

$$\frac{1}{2+4/(n+1)}\Big[-\frac{n+1}{2}log(1+2/(n+1))\Big] \leq \frac{1}{2+4/n}\Big[-\frac{n}{2}log(1+2/n)\Big].$$

Changing variables to $x := (n+1)/2$ and $y := n/2$ leads to

$$\frac{xlog(1+1/x)}{1+1/x} \geq \frac{ylog(1+1/y)}{1+1/y}$$

and changing variables again $\alpha := 1+1/x > 1$ and $\beta := 1+1/y > 1$ we remain with

$$\frac{log(\alpha)}{\alpha(\alpha-1)} \geq \frac{log(\beta)}{\beta(\beta-1)}.$$

So now it remains to show that the function $f : \mathbb{R} \to \mathbb{R}$, defined by

$$f(t) = \frac{log(t)}{t(t-1)},$$

is decreasing in t and this would lead to the conclusion since $\alpha < \beta$. Computing its derivative $f'(t)$ one gets:

$$f'(t) = \frac{t-1-log(t)(2t-1)}{t^2(t-1)^2}.$$

We have to verify the inequality just for $t \geq 1$. We define then

$$g(t) = log(t) - \frac{t-1}{2t-1}$$

and compute its derivative:

$$g'(t) = \frac{(2t-1)^2 - t}{t(2t-1)^2}$$

and so we can see (remember $t \geq 1$) that $g'(t) \leq 0$ if and only if $t \leq 1$, and $g'(1) = 0$, so $t = 1$ is a minimum. $g(1) = 0$ and then positive. So, going backwards with the computations, the inequality $S_h(n+1) < S_h(n)$ is verified.
Now we have to prove the asymptotics and this is easy:

$$\lim_{n\to+\infty} S(n) = \lim_{n\to+\infty} \Big(\frac{1}{2}(1+\frac{2}{n})^{-n/2}\Big)^{\frac{1}{2+4/n}} = \lim_{n\to+\infty} 2^{-1/2} 1/e^{\frac{1}{2+4/n}} = \frac{1}{\sqrt{2e}}.$$

It remains to prove the equivalence between the homogeneous and the dual constant. It basically comes from a duality argument. Denote with $< \cdot, \cdot >$ the dual product (it changes accordingly to the space) and define $Tu := e^{it\Delta}u$. Then for every $f \in L^2_x$ an $F \in L^q_t L^r_x$ we have

$$| < f, T^*F > | = | < Tf, F > | \leq ||Tf||_{L^q_t L^r_x} ||F||_{L^{q'}_t L^{r'}_x} \leq S_h ||f||_{L^2_x} ||F||_{L^{q'}_t L^{r'}_x}.$$

So

$$||T^*F||_{L^2_x} := \sup_{f\in L^2_x} \frac{|<f,T^*F>|}{||f||_{L^2_x}} \le S_h||F||_{L^{q'}_t L^{r'}_x},$$

hence $S_d \le S_h$. Analogously:

$$|<Tf,F>| = |<f,T^*F>| \le ||f||_{L^2_x}||T^*F||_{L^2_x} \le S_d||f||_{L^2_x}||F||_{L^{q'}_t L^{r'}_x}.$$

So

$$||Tf||_{L^q_t L^r_x} := \sup_{F\in L^{q'}_t L^{r'}_x} \frac{|<Tf,F>|}{||F||_{L^{q'}_t L^{r'}_x}} \le S_d||f||_{L^2_x},$$

hence $S_h \le S_d$ and so we get $S_h = S_d$. This concludes the proof of the theorem.

1.2.4 Proof of Theorem 1.2.5

Here, we give the proof of Theorem 1.2.5. We will skip some of the details because standard in the theory of global wellposedness for the NLS. We refer to [116] for some of the details skipped. We consider equation (1.1):

$$i\partial_t u(t,x) + \Delta u(t,x) + \mu|u|^{p-1}u(t,x) = 0 \quad (t,x)\in(0,\infty)\times\mathbb{R}^n, \tag{1.14}$$

with initial datum $u(0,x) = u_0(x)$, space dimension is $n \ge 1$, $p \ge 1$ in both the focusing and defocusing case: $\mu = -1, 1$, since we are dealing with a small data analysis. By Duhamel's Principle we define the Duhamel's Operator:

$$Lu := \chi(t/T)e^{-it\Delta}u_0(x) - i\mu\chi(t/T)\int_0^t e^{-i(t-s)\Delta}|u(s,x)|^{p-1}u(s,x)ds, \tag{1.15}$$

where $T > 0$ and $\chi(r)$ is a smooth cut-off function supported on $-2 \le r \le 2$ and such that $\chi(r) = 1$ on $-1 \le r \le 1$. Using Duhamel's formula, we take the $L^q_t L^r_x$-norm of Lu (from now on, unless specified, $t \in [-T,T]$ in the definition of $L^q_t L^r_x$-norm), and get:

$$||Lu||_{L^q_t L^r_x} \le S_h(n,r)||u||_{L^2_x} + S_i(n,r)||u||^p_{L^{\tilde{q}'p}_t L^r_x}$$

$$\le S_h(n,r)||u||_{L^2_x} + S_i(n,r)T^{1/(\tilde{q}')-p/q}||u||^p_{L^q_t L^r_x}$$

choosing $\tilde{r}'p = r$.

Now we need to do some numerology. Since (q,r) and $(\tilde{q},\tilde{r})$ are admissible pairs: $2/q + n/r = n/2$, $2/\tilde{q} + n/\tilde{r} = n/2$. Moreover, since we are in the L^2-critical case we can choose $\tilde{r}'p = r$ and $\tilde{q}'p = q$, having still some freedom on the choice of (q,r) as it can be seen by the following lemma. The conditions on (q, r) and $(\tilde{q}, \tilde{r})$ can be rewritten as a system of linear equations in $(1/q, 1/\tilde{q}, 1/r, 1/\tilde{r})$.

Lemma 1.2.13. *There exist infinitely many solutions to the system*

$$SE = N,$$

where

$$S = \begin{pmatrix} 2 & 0 & n & 0 \\ 0 & 2 & 0 & n \\ 0 & 0 & p & 1 \\ p & 1 & 0 & 0 \end{pmatrix},$$

$E = (1/q, 1/\tilde{q}, 1/r, 1/\tilde{r})^T$ *and* $N = (n/2, n/2, 1, 1)^T$, *if and only if* $p = 1 + 4/n$. *If* $p \neq 1 + 4/n$ *the system has no solutions.*

Remark 1.2.14. *Basically this lemma implies that, using the estimates that we have used above in the* H^s*-scale, we cannot remove a power of* T *in front of the nonlinear term in the* subcritical *(good) and* supercritical *(bad) cases.*

Proof. We can see that $det(S) = 0$ and $rank(S) = 3$, because the upper-left 3×3 matrix is not singular for $p \neq 0$. If $p \neq 1+4/n$, then the augmented matrix $[S, N]$ has $rank([S, N]) = 4$, so the system has no solutions, while for $p = 1+4/n$, $rank([S, N]) = 3$ and so the system has infinitely many solutions. □

Remark 1.2.15. *Similar computations can be done for any regularity* s*, and with nonlinear exponent* $p(s) = 1 + 4/(n - 2s)$*. The* critical *case* $\tilde{q}'p = q$ *is the interesting one for us, because in the* subcritical *case* $\tilde{q}'p < q$*, we can shrink the interval of local wellposedness, since* $T^{1/(\tilde{q}') - p/q}$ *appears in the estimates of Duhamel's Operator with a positive power, and so we do not need to do a small data theory.*

Now we will see how big the datum can be in order to have a "cheap" contraction with only the estimates done above. Define

$$R := \alpha S_h(n, r) ||u_0||_{L^2_x}$$

and

$$B_R := \{u \in L^q_t L^r_x : ||u||_{L^q_t L^r_x} \leq R\}.$$

Choose also $\beta > 0$ such that

$$S_i(n, r) R^{p-1} < 1/\beta.$$

With these choices we get:

$$||Lu||_{L_t^q L_x^r} \leq S_h(n,r)||u||_{L_x^2} + S_i(n,r)T^{1/(\tilde{q}')-p/q}||u||^p_{L_t^q L_x^r} \leq R(1/\alpha + 1/\beta) \leq R$$

for every $1/\alpha + 1/\beta \leq 1$ and with $1/\alpha + 1/\beta = 1$ in the less restrictive case. So the Duhamel's Operator L sends the balls B_R into themselves if $||u_0||_{L_x^2}$ is small enough, more precisely when

$$||u_0||_{L_x^2} = \frac{R}{S_h(n)\alpha}.$$

This implies that

$$S_i(n,r)\left(\alpha S_h(n,r)||u||_{L_x^2}\right)^{p-1} < 1/\beta,$$

which means:

$$||u||_{L_x^2} < \frac{1}{S_h(n,r)\alpha}\left(\frac{1}{\beta S_i(n,r)}\right)^{1/(p-1)}.$$

Using our hypotheses on p, α, β we get:

$$||u||_{L_x^2} < \frac{1}{S_h(n,r)\alpha}\left(\frac{1}{S_i(n,r)} - \frac{1}{S_i(n,r)\alpha}\right)^{n/4}. \tag{1.16}$$

For now, the only restriction on α is $1/\alpha + 1/\beta \leq 1$.

Remark 1.2.16. *The coefficients α and β are almost conjugate exponents, suggesting an orthogonal decomposition of the solution on the linear flow and on the nonlinear one.*

Now we check that the operator Lu is a contraction. Let

$$u(t) = e^{-it\Delta}u_0 - i\mu\int_0^t e^{-i(t-s)\Delta}|u(s)|^{p-1}u(s)ds. \tag{1.17}$$

and

$$v(t) = e^{-it\Delta}u_0 - i\mu\int_0^t e^{-i(t-s)\Delta}|v(s)|^{p-1}v(s)ds. \tag{1.18}$$

be two solutions of (1.14). Then

$$\begin{aligned}||Lu - Lv||_{L_t^q L_x^r} &= \left\| \int_0^t e^{-i(t-s)\Delta} \left(|u(s)|^{p-1} u(s) - |v(s)|^{p-1} v(s) \right) ds \right\|_{L_t^q L_x^r} \\ &\leq S_i(n,r) ||\,|u|^{p-1}u - |v|^{p-1}v||_{L_t^{\tilde{q}'} L_x^{\tilde{r}'}} \\ &\leq S_i(n,r) \Big(||u||^{p-1}_{L_t^q L_x^r} + ||v||^{p-1}_{L_t^q L_x^r} \Big) ||u - v||_{L_t^q L_x^r}\end{aligned}$$

in the above choice of exponents (q,r) and $(\tilde{q},\tilde{r})$. This implies:

$$||Lu - Lv||_{L_t^q L_x^r} \leq 2S_i(n) R^{p-1} ||u - v||_{L_t^q L_x^r} < 2/\beta ||u - v||_{L_t^q L_x^r},$$

so we need $2/\beta \leq 1$, namely $\beta \geq 2$ and so $1 \leq \alpha \leq 2$, since $1/\alpha + 1/\beta \leq 1$. This is the last restriction on α that we need to apply to the estimate (1.16). We remark here that (1.16) holds for every $1 \leq \alpha \leq 2$ and so we are allowed to take the maximum on both sides of (1.16). Notice also that the left hand side of (1.16) does not depend on α.

Remark 1.2.17. *To have a contraction the ball needs to be big enough, but not that much namely $S_h(n,r)||u||_{L_x^2} \leq R \leq 2S_h(n,r)||u||_{L_x^2}$.*

Now we want to optimize on $||u_0||_{L_x^2}$, namely we want to take it as big as possible, maintaining the property of Lu of being a contraction. In other words we have to find the maximum of the function

$$F_n(\alpha) = \frac{1}{\alpha} \left(1 - \frac{1}{\alpha} \right)^{n/4},$$

when $\alpha \in [1,2]$. Taking the derivative, we get:

$$F_n'(\alpha) = -\alpha^{-2-n/4} (\alpha - 1)^{n/4-1} \left(-(1 + n/4)(\alpha - 1) + \alpha n/4 \right).$$

So $F_n'(\alpha) \geq 0$ if and only if

$$1 \leq \alpha \leq 1 + n/4.$$

In particular when $n \geq 4$, $\alpha_{max} = 2$ and when $n \leq 4$, $\alpha_{max} = 1 + n/4$. This concludes the proof of Theorem 1.2.5.

Remark 1.2.18. *The coefficient $\alpha = 2$ is not always the optimal one, as it is usually used in every exposition on the topic. The optimal α depends on the dimension n. We can compute explicitly the values of $F_n(\alpha_{max})$ in any dimension: for $n = 1$ $F_n(\alpha_{max}) = F_1(5/4) = 5^{-5/4}4$, for $n = 2$ $F_n(\alpha_{max}) = F_2(3/2) = 3^{-3/2}2$, for $n = 3$ $F_n(\alpha_{max}) = F_3(7/4) = 3^{3/4}7^{-7/4}4$ and for $n \geq 4$ $F_n(\alpha_{max}) = 2^{-1-n/4}$.*

1.2.5 Applications to Fourier Restriction Inequalities

Strichartz Inequalities can be set in the more general framework of Fourier Restriction Inequalities in Harmonic Analysis. This connection has been made clear already in the original paper of Strichartz [113]. In this subsection we will highlight this relationship in the Schrödinger/Paraboloid case and we will see how to prove Theorem 1.2.7. For the case of different flows and hypersurfaces, like the Wave/Cone or Helmholtz/Sphere cases, we refer to [117] and the references therein for more details.
Consider a function $f \in L^1(\mathbb{R}^n)$, then its Fourier Transform $\hat{f}$ is a bounded and continuous function on all $\mathbb{R}^n$ and it vanishes at infinity. So $\hat{f}|_{\mathcal{S}}$, the restriction of $\hat{f}$ to a set $\mathcal{S}$ is well defined even if $\mathcal{S}$ has measure zero, like, for example, if $\mathcal{S}$ is a hypersurface. It becomes then interesting to understand what happens if $f \in L^p(\mathbb{R}^n)$ for $1 < p < 2$. From Hausdorff-Young inequality, we can see that if $f \in L^p(\mathbb{R}^n)$ then $\hat{f} \in L^{p'}(\mathbb{R}^n)$ with $1/p + 1/p' = 1$, so $\hat{f}$ can be naturally restricted to any set $\mathcal{A}$ of positive measure. It turns out that a big role is played by the geometry of the set $\mathcal{S}$. Stein proved that if the set $\mathcal{S}$ is sufficiently smooth and its curvature is big enough (in fact it is not true for hyperplanes), then it makes sense to talk about $\hat{f}|_{\mathcal{S}}$ belonging to L^p-spaces.

1.2.5.1 Proof of Theorem 1.2.7

From now on we will focus on the case where the hypersurface is the paraboloid $\mathcal{S} = \mathbf{P}^n$, where $\mathbf{P}^n$ is defined as

$$\mathbf{P}^n := \{(\tau, \xi) \in \mathbb{R} \times \mathbb{R}^n : -\tau = |\xi|^2\} \tag{1.19}$$

and is endowed with the measure dP^n that is given by

$$\int_{\mathbf{P}^n} h(\tau, \xi) dP^n = \int_{\mathbb{R}^n} h(-|\xi|^2, \xi) d\xi. \tag{1.20}$$

(here h is a Schwartz function) and induced by the embedding $\mathbf{P}^n \hookrightarrow \mathbb{R}^{n+1}$. To prove the theorem, we have just to show the equivalence of Restriction Inequalities and Strichartz Inequalities.
It makes sense to talk about a restriction, if $\hat{f}|_{\mathcal{S}}$ is not infinite almost everywhere and a *restriction estimate* holds:

$$||\hat{f}|_{\mathbf{P}^n}||_{L^q(\mathbf{P}^n, dP^n)} \leq ||f||_{L^p(\mathbb{R}^n)},$$

for some $1 \leq q < \infty$ and for every Schwartz function f. This last estimate is equivalent, by a duality argument and Parseval Identity, to

$$||\mathcal{F}^{-1}(\hat{F}dP^n)|_{\mathbf{P}^n}||_{L^{p'}(\mathbb{R}^n)} \leq ||F||_{L^{q'}(\mathbf{P}^n, dP^n)},$$

for all Schwartz functions F on $\mathbf{P}^n$ and where

$$\mathcal{F}^{-1}(\hat{F}dP^n)(t,x) = \int_{\mathbf{P}^n} e^{ix\xi + it\tau}\hat{F}(\tau,\xi)d\tau dP^n(\tau,\xi)$$

is the Inverse Space-Time Fourier Transform of the measure $\hat{F}dP^n$. The dual formulation connects directly to the fundamental solution (1.10)

$$u(t,x) = (1-4it)^{-n/2}e^{-\frac{|x|^2}{1-4it}}$$

of equation (1.9)

$$i\partial_t u(t,x) = \Delta u(t,x), \quad (t,x) \in (0,\infty) \times \mathbb{R}^n.$$

Since u can be rewritten in the form

$$u = \mathcal{F}^{-1}(\hat{u}_0 dP^n).$$

In this way the *Homogeneous Strichartz Inequality*

$$\left|\left|e^{it\Delta}u_0\right|\right|_{L^q_t L^r_x} \leq S_h(n,q,r)||u_0||_{L^2_x},$$

for $q = r = 2 + 4/n$, as in this present case, can be rewritten in the following way:

$$||\widehat{fdP^n}||_{L^{\frac{2(n+2)}{n}}_{t,x}(\mathbb{R}^{n+1})} \leq S_h(n)||f||_{L^2(\mathbf{P}^n, dP^n)} \tag{1.21}$$

with $S_h(n)$ given by

$$S_h(n) = \left(\frac{1}{2}(1+\frac{2}{n})^{-n/2}\right)^{\frac{1}{2+4/n}}.$$

This proves Theorem 1.2.7.

Remark 1.2.19. *We notice that results for the paraboloid seem easier to obtain than for example for the sphere. For example there is not yet the counterpart of [27] in the wave/sphere case and we do not have a conjecture on the sharp Strichartz Constant in general dimension in the case of the wave equation.*

Remark 1.2.20. *As we said above, the connection between restriction theorems and PDE links a much broader class of hypersurfaces and PDEs. For more details on the more* *recent results, we refer to [27], [28], [29], [48], [49] and to [117] for a survey on restriction theorems.*

Remark 1.2.21. *The Hilbert structure has been crucial in some of the proofs of the existence of maximizers for restriction inequalities. See for example [48] and [27]. Here, we are in L^2_x and so a Hilbert case, but our analysis is not touched by this problem, because we are interested in the optimal constants and not on the extremizers.*

1.2.6 Appendix: some comments on the Inhomogeneous Case and the Wave Equation

In this appendix, we share some comments and computations on the Inhomogeneous Strichartz Estimate and on the case of the wave equation. We will not prove any theorem, but we will highlight some difficulties and make some remarks.

1.2.6.1 The Inhomogeneous Strichartz constant S_i

By the TT^* principle (take $Tu := e^{it\Delta}$) and by duality, the Homogeneous Strichartz and the Dual Strichartz inequality are equivalent. By the same principle one can prove that the operator $TT^* : L^q_t L^r_x \to L^{\tilde{q}'}_r L^{\tilde{r}'}_x$ is bounded if and only if the operator $T : L^2_x \to L^q_t L^r_x$ is bounded. Unfortunately, the Inhomogeneous Strichartz Inequality cannot be seen as such a composition because it involves the retarded operator. This does not prevent the retarded operator to keep the boundedness properties of TT^* but it complicates a lot the computation of $S_i(n, r, q, \tilde{r}, \tilde{q})$ and the proof of the existence of critical points, that, as far as we know, has not been treated yet in the literature. In the following, we will outline how the integrals become not tractable in the inhomogeneous case already in the case of a Gaussian and so a simple direct computation seems not to be enough to calculate the best Strichartz Constant. We will concentrate also here on the L^2-critical case. See [116] or [70] for more details on the TT^*-method.
We now test the inhomogeneous inequality with Gaussians for every dimension. It is not known yet in the literature if they are maximizers or not, but an explicit computation would lead at least to a lower bound on the constant. We recall that the solutions that we want to test are

$$u(t, x) = (1 - 4it)^{-n/2} e^{-\frac{|x|^2}{1-4it}},$$

while the inequality we need to test is

$$\left\| \int_{s<t} e^{-i(t-s)\Delta} F(s) ds \right\|_{L^q_t L^r_x} \leq S_i(n, q, r\tilde{q}, \tilde{r}) ||F||_{L^{\tilde{q}'}_t L^{\tilde{r}'}_x}.$$

with $F(t,x) = |u(t,x)|^{p-1}u(t,x)$.
We start by computing the norm on the right hand side of this inequality. By the choice of the exponents and the criticality of the problem $\tilde{r}'p = r$ and $\tilde{q}'p = q$. So we get

$$||F||_{L_t^{\tilde{q}'} L_x^{\tilde{r}'}} = |||u|^p||_{L_t^{q/p} L_x^{r/p}} = ||u||^p_{L_t^q L_x^r}.$$

By the computations done in Section 1.2.3, we then get:

$$||F||_{L_t^{\tilde{q}'} L_x^{\tilde{r}'}} = \left(\frac{\pi}{4}\right)^{\frac{np(r-2)}{4r}} \left(\frac{\pi}{r}\right)^{\frac{pn}{2r}}.$$

Now we have to compute the left hand side of the inhomogeneous Strichartz Inequality:

$$\left\| \int_{s<t} e^{-i(t-s)\Delta} F(s) ds \right\|_{L_t^q L_x^r}.$$

We start computing explicitly $e^{-i(t-s)\Delta}F(s)$. By definition of $e^{-i(t-s)\Delta}$, we have

$$e^{-i(t-s)\Delta}F(s) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \widehat{e^{-i(t-s)\Delta} F} d\xi = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{ix\cdot\xi + i(t-s)|\xi|^2} \hat{F} d\xi.$$

So we have now to compute $\hat{F}(s,\xi)$:

$$\begin{aligned}
\hat{F}(s,\xi) &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} F(s,x) dx = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} |u(s,x)|^{p-1} u(s,x) dx \\
&= (2\pi)^{-n/2} \left|1+16s^2\right|^{-(p-1)n/4} (1-4is)^{-n/2} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} e^{-\frac{|x|^2}{1-4is} - \frac{(p-1)|x|^2}{1+16s^2}} \\
&= (2\pi)^{-n/2} \left|1+16s^2\right|^{-(p-1)n/4-n/2} (1+4is)^{n/2} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} e^{-\frac{(p+4is)|x|^2}{1+16s^2}} \\
&= (2\pi)^{-n/2} \left|1+16s^2\right|^{-(p-1)n/4-n/2} (1+4is)^{n/2} \times \\
&\times \int_{\mathbb{R}^n} e^{-\frac{(p+4is)|x|^2}{1+16s^2} - ix\cdot\xi + \frac{(1+16s^2)|\xi|^2}{4(p+4is)}} e^{-\frac{(1+16s^2)|\xi|^2}{4(p+4is)}} \\
&= (2\pi)^{-n/2} \left|1+16s^2\right|^{-(p-1)n/4-n/2} (1+4is)^{n/2} \times \\
&\times \int_{\mathbb{R}^n} e^{-\left|\frac{x(p+4is)^{1/2}}{(1+16s^2)^{1/2}} + i\frac{\xi(1+16s^2)^{1/2}}{2(p+4is)^{1/2}}\right|^2} e^{-\frac{(1+16s^2)|\xi|^2}{4(p+4is)}} \\
&= \frac{e^{-\frac{(1+16s^2)|\xi|^2}{4(p+4is)}}}{(2\pi)^{n/2}} \\
&\times \left|1+16s^2\right|^{-(p-1)n/4-n/2} (1+4is)^{n/2} (1+16s^2)^{n/2} (p+4is)^{-n/2} \pi^{n/2} \\
&= 2^{-n/2} \left|1+16s^2\right|^{-(p-1)n/4} \left(\frac{1+4is}{p+4is}\right)^{n/2} e^{-\frac{(1+16s^2)|\xi|^2}{4(p+4is)}}
\end{aligned}$$

by completing the square and changing integration variables to

$$y=\frac{x(p+4is)^{1/2}}{(1+16s^2)^{1/2}}+i\frac{\xi(1+16s^2)^{1/2}}{2(p+4is)^{1/2}},$$

similarly to the computations that we have done in Section 1.2.2. So

$$\hat{F}(s,\xi)=2^{-n/2}\left|1+16s^2\right|^{-(p-1)n/4}\left(\frac{1+4is}{p+4is}\right)^{n/2}e^{-\frac{(1+16s^2)|\xi|^2}{4(p+4is)}}.$$

Notice that this is consistent with what we got in Section 1.2.2 in the case $s=0$ and $p=1$. Now, putting everything together, we get:

$$\begin{aligned}
&e^{-i(t-s)\Delta}F(s)=\frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n}e^{ix\cdot\xi+i(t-s)|\xi|^2}\hat{F}d\xi=\\
&\frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n}e^{ix\cdot\xi+i(t-s)|\xi|^2}2^{-n/2}\left|1+16s^2\right|^{-(p-1)n/4}\left(\frac{1+4is}{p+4is}\right)^{n/2}e^{-\frac{(1+16s^2)|\xi|^2}{4(p+4is)}}=\\
&\frac{1}{(2\pi)^{\frac{n}{2}}}2^{-n/2}\left|1+16s^2\right|^{-(p-1)n/4}\left(\frac{1+4is}{p+4is}\right)^{n/2}\int_{\mathbb{R}^n}e^{-\frac{(1+16s^2)|\xi|^2}{4(p+4is)}}e^{ix\cdot\xi+i(t-s)|\xi|^2}=\\
&\frac{1}{(2\pi)^{\frac{n}{2}}}2^{-n/2}\left|1+16s^2\right|^{-(p-1)n/4}\left(\frac{1+4is}{p+4is}\right)^{n/2}\times\\
&\times\int_{\mathbb{R}^n}e^{-|\xi|^2\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]+ix\xi+\frac{|x|^2}{4\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]}}e^{-\frac{|x|^2}{4\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]}}=\\
&\frac{1}{(2\pi)^{\frac{n}{2}}}2^{-n/2}\left|1+16s^2\right|^{-(p-1)n/4}\left(\frac{1+4is}{p+4is}\right)^{n/2}\times\\
&\times\int_{\mathbb{R}^n}e^{-\left|\xi\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]^{1/2}-\frac{ix}{2\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]^{1/2}}\right|^2}e^{-\frac{|x|^2}{4\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]}}
\end{aligned}$$

which by the change of variable $\eta=\xi\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]^{1/2}-\frac{ix}{2\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]^{1/2}}$, becomes

$$\begin{aligned}
e^{-i(t-s)\Delta}F(s)=&\frac{1}{(2\pi)^{\frac{n}{2}}}2^{-n/2}\left|1+16s^2\right|^{-(p-1)n/4}\left(\frac{1+4is}{p+4is}\right)^{n/2}\times\\
&\times\int_{\mathbb{R}^n}e^{-|\eta|^2}e^{-\frac{|x|^2}{4\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]}}\left[\frac{(1+16s^2)}{4(p+4is)}-i(t-s)\right]^{-n/2}.
\end{aligned}$$

In conclusion:

$$
\begin{aligned}
&e^{-i(t-s)\Delta}F(s) = \frac{1}{(2\pi)^{\frac{n}{2}}} 2^{-n/2} \left|1+16s^2\right|^{-(p-1)n/4} \left(\frac{1+4is}{p+4is}\right)^{n/2} \\
\times \quad &\left[\frac{(1+16s^2)}{4(p+4is)} - i(t-s)\right]^{-n/2} \times \\
&\times \pi^{n/2} e^{-\frac{|x|^2}{4\left[\frac{(1+16s^2)}{4(p+4is)} - i(t-s)\right]}} = \\
&\left|1+16s^2\right|^{-(p-1)n/4} \left(\frac{1+4is}{p+4is}\right)^{n/2} \left[\frac{(1+16s^2)}{p+4is} - 4i(t-s)\right]^{-n/2} e^{-\frac{|x|^2}{4\left[\frac{(1+16s^2)}{4(p+4is)} - i(t-s)\right]}} = \\
&\left|1+16s^2\right|^{-(p-1)n/4} \left[\frac{1+4is}{1-4ip(t-s)+16ts}\right]^{\frac{n}{2}}
\end{aligned}
$$

Again, this is consistent with what we got in Section 1.2.2 in the case $s = t = 0$ and $p = 1$. At this point the approach of the direct computation seems not good enough anymore, because one should integrate in the variable s and this does not seem to have an explicit expression with elementary functions. We plan to study numerically this case in a future work.

Remark 1.2.22. *If one would be able to compute explicitly $S_i(n, r)$, one could use Theorem 1.2.5 also as a stability-type result for the solutions of the NLS, in a similar spirit of the stability of solitons in the focusing case. This connection links, in some sense, optimizers and stability, also when the functionals involve both space and time.*

1.2.6.2 The wave Equation case

For completeness, we want to mention here that similar studies have been done for several others homogeneous Strichartz Estimates, like the wave equation. The complete characterization of critical points done by [27] in the case of the Schrödinger Equation is still not available in the case of the wave equation. We believe that an argument completely similar to the one that we have given in Section 1.2.3 would lead to the computation of the possible best Homogeneous Wave Strichartz Constant $W(n)$ for the wave equation, once a complete characterization of the maximizers would be available. For more details on the case of the wave equation we refer to [53], [24] and [21].

Remark 1.2.23. *There are well known transformations that send solutions to the Schrödinger Equation to solutions of the wave equation, see for example [116]. So one strategy here could be also to transform the maximizers of $S_h(n, r)$ into solutions of the corresponding wave equation and hope that the known transformation sends maximizers to* *maximizers. Unfortunately, to our knowledge, no known transformation does this job. This technique could be very helpful also for other equations.*

Remark 1.2.24. *We mention here that the functions that optimize the Wave Strichartz Inequality (see [53]), optimize also the Sobolev Embeddings (see [115], [5] and [6]). Let* $1 < p < n$ *and* $p^* = \frac{np}{n-p}$*, then*

$$||u||_{L^{p^*}(\mathbb{R}^n)} \leq C(n,p)||\nabla u||_{L^p(\mathbb{R}^n)}$$

with optimal constant $C(n,p)$ *given by*

$$C(n,p) = \pi^{1/2} n^{-1/p} \left(\frac{p-1}{n-p}\right)^{1-1/p} \left(\frac{\Gamma(1+n/2)\Gamma(n)}{\Gamma(n/p)\Gamma(1+n-n/p)}\right)$$

and maximizers given by

$$u(x) = (a + b|x|^{\frac{p}{p-1}})^{-\frac{n-p}{p}},$$

with $a, b > 0$*. We notice that with* $p = 2$ *and substituting* n *with* $n+1$ *in the above optimizers, we recover the optimizers given in [53]. The correspondence between the constants seems more involved.*

1.3 The Maximal Strichartz Family of Gaussian Distributions: Fisher Information, Index of Dispersion, Stochastic Ordering

In this section, we define and study several properties of what we call *Maximal Strichartz Family of Gaussian Distributions*. This is a subfamily of the family of Gaussian Distributions that arises naturally in the context of the *Linear Schrödinger Equation* and Harmonic Analysis, as the set of maximizers of certain norms introduced by Strichartz, as discussed in the previous section of this chapter.

From a statistical perspective, this family carries with itself some extra-structure with respect to the general family of Gaussian Distributions. In this section, we analyse this extra-structure in several ways. We first compute the *Fisher Information Matrix* of the family, we then introduce some measures of *Statistical Dispersion* and, finally, we introduce a *Partial Stochastic Order* on the family. Moreover, we indicate how these tools can be used to distinguish between distributions which belong to the family and distributions which do not. We show also that all our results are in accordance with the Dispersive-PDE nature of the family.

1.3.1 Introduction and Motivation

The most important multivariate distribution is the *Multivariate Normal Distribution* (MVN). To fix the notation, we give here its definition.

Definition 1.3.1. *We say that a random variable X is distributed as a* Multivariate Normal Distribution *if its* probability density function *(pdf)*

$$f_X : \mathbb{R}^n \to \mathbb{R}$$

takes the form

$$f_X(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right),$$

where $\mu := E[X] \in \mathbb{R}^n$ is the Mean Value Vector *and $\Sigma := Var(X) \in Sym_n^+$ is the $n \times n$ positive definite symmetric* Variance-Covariance Matrix.

Its importance derives mainly (but not only) from the *Multivariate Central Limit Theorem* which has the following statement.

Theorem 1.3.2. *Suppose that $X = (x_1, \dots, x_n)^T$ is a random vector with Variance-Covariance Matrix Σ. Assume also that $E[x_i^2] < +\infty$ for every $i = 1, \dots, n$. If $X_1, X_2, \dots$ is a sequence of iid random variables distributed as X, then*

$$\frac{1}{n^{1/2}} \Sigma_{i=1}^n \Big(X_i - E[X_i] \Big) \to^d MVN(0, \Sigma),$$

where $\to^d$ represents the convergence in distribution.

Due to its importance, several authors have tried to give characterizations of this family of distributions. See for example [90] and [72] for an extended discussion on multivariate distributions and their properties. As mentioned in Section 1.1 the MVN can be characterized by means of variational principles, such as the maximization of certain functionals. We recall from Section 1.1, the well known characterization of the Gaussian distribution through the maximization of the *Differential Entropy*, under the constraint of fixed variance Σ in the case when the support of the pdf is the whole Euclidean Space $\mathbb{R}^n$.

Theorem 1.3.3. *Let X be a random variable whose pdf is f_X. The* Differential Entropy *$h(X)$ is defined by the following functional:*

$$h(X) := -\int_{\mathbb{R}^n} f_X(x) \log f_X(x)\, dx.$$

The Multivariate Normal Distribution *has the largest* Differential Entropy *$h(X)$ amongst all the random variables X with equal variance Σ. Moreover, the maximal value of the* Differential Entropy *$h(X)$ is $h(MVN(\Sigma)) = \frac{1}{2}log[(2\pi e)^n|\Sigma|$.*

We refer to Section 1.1 for a proof of this well known theorem. This characterization is, in some sense, not completely satisfactory because it is given just with the restriction of fixed variance.

As mentioned in Section 1.2, a more general characterization of the *Gaussian Distribution* has been given in a setting which, at first sight, seem very far, and it is the one of Harmonic Analysis and Partial Differential Equations.

We first introduce the so called *Admissible Exponents*.

Definition 1.3.4. *Fix $n \geq 1$. We call a set of exponents (q, r)* admissible *if $2 \leq q, r \leq +\infty$ and*

$$\frac{2}{q} + \frac{n}{r} = \frac{n}{2}.$$

Remark 1.3.5. *As mentioned in Section 1.2, these exponents are characteristic quantities of certain norms, the* Strichartz Norms*, naturally arising in the context of Dispersive Equations and can vary from equation to equation. We refer to [116] for more details.*

We recall the precise characterization of the Multivariate Normal Distribution, through Strichartz Estimates, towards which we contributed in Theorem 1.2.1 of previous section.

Theorem 1.3.6. *[113], [68], [27], Suppose $n = 1$ or $n = 2$. Then, for every (q, r) and $(\tilde{q}, \tilde{r})$ admissible and for every $u_0 \in L^2_x(\mathbb{R}^n)$ such that $||u_0||^2_{L^2(\mathbb{R}^n)} = 1$, we have*

$$\left|\left|e^{-it\Delta} u_0\right|\right|_{L^q_t L^r_x} \leq S(n, q, r), \tag{1.22}$$

where $S_h(n, q, r) = S_h(n, r)$ is the sharp Homogeneous Strichartz Constant*, defined by*

$$S_h(n, r) := \sup\left\{||u||_{L^q_t L^r_x(\mathbb{R}\times\mathbb{R}^n)} : ||u||^2_{L^2_x(\mathbb{R}^n)} = 1\right\}, \tag{1.23}$$

and given by

$$S_h(n,r) = 2^{\frac{n}{4}-\frac{n(r-2)}{2r}} r^{-\frac{n}{2r}}. \tag{1.24}$$

Moreover, the inequality (1.22) becomes an equality if and only if $|u_0|^2$ *is the pdf of a* Multivariate Normal Distribution.

For several other important results on Strichartz estimates we refer to [9], [63], [65], [71] and the references therein.

The symmetries of the functional in (1.22) give rise to a family of distributions that we call *Maximal Strichartz Family of Gaussian Distributions.*

$$\mathcal{F} := \Big\{ p(t,x) = \left(\frac{\pi}{2}\right)^{-\frac{n}{2}} |R^T R|^{-\frac{1}{2}} |\lambda^2 + 16t^2|^{-n/2} e^{-\frac{2(x-x_0-vt)^T (R^T R)^{-1}(x-x_0-vt)}{\lambda^2+16t^2}} \tag{1.25}$$

$$(t,\lambda) \in \mathbb{R}\times\mathbb{R}, (x_0,v)\in\mathbb{R}^n\times\mathbb{R}^n, R\in SO(n) \Big\}. \tag{1.26}$$

We refer to Section 1.3.2 for its precise construction. This is a subfamily of the family of Gaussian Distributions and, among the other things, it has the feature that the *Mean Vector* μ and the *Variance Covariance Matrix* Σ depend on common parameters. Therefore, from a statistical perspective, this family carries with itself some extra-structure with respect to the general family of Gaussian Distributions. This extra-structure becomes evident from the form of the *Fisher Information Metric* of the family.

Theorem 1.3.7. *Consider* $p(t,x)$*, a probability distribution function belonging to the* Maximal Strichartz Family of Gaussian Distributions $\mathcal{F}$*, defined in equation (1.25). The vector of parameters* θ*, indexing* $\mathcal{F}$*, is given by*

$$\theta := (x_0^T, v_0^T, \lambda, (R^T R)_{ij}, t)^T.$$

Then, the Fisher Information Matrix *of* $p(t,x)$ *is given by the following:*

- *in the spherical case* ($R^T R = \sigma^2 Id$)*, by*

$$
I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} Id & \frac{t}{\sigma^2} Id & 0 & 0 & \frac{1}{\sigma^2} v_0 \\ \frac{t}{\sigma^2} Id & \frac{t^2}{\sigma^2} Id & 0 & 0 & \frac{t}{\sigma^2} v_0 \\ 0 & 0 & \frac{\lambda^2 n}{8} & n\frac{\lambda}{16\sigma^2}(\lambda^2+16t^2) & 2\lambda t n \\ 0 & 0 & n\frac{\lambda}{16\sigma^2}(\lambda^2+16t^2) & \frac{n}{32}\frac{\left(\lambda^2+16t^2\right)^2}{\sigma^4} & \frac{nt}{\sigma^2}\left(\lambda^2+16t^2\right) \\ \frac{1}{\sigma^2} v_0 & \frac{t}{\sigma^2} v_0 & 2\lambda t n & \frac{nt}{\sigma^2}\left(\lambda^2+16t^2\right) & \frac{|v_0|^2}{\sigma^2} + 32nt^2 \end{pmatrix};
$$

- *in the elliptical case ($R^T R = \sigma_i^2 Id$ -see Section 1.3.3 for the precise definition), by*

$$
I(\theta) = \begin{pmatrix} \frac{1}{\sigma_i^2} Id & \frac{t}{\sigma_i^2} Id & 0 & 0 & \frac{1}{\sigma_i^2} v_0^i \\ \frac{t}{\sigma_i^2} Id & \frac{t^2}{\sigma_i^2} Id & 0 & 0 & \frac{t}{\sigma_i^2} v_0^i \\ 0 & 0 & \frac{\lambda^2 n}{8} & \frac{\lambda}{16\sigma^2}(\lambda^2+16t^2) & 2\lambda t n \\ 0 & 0 & \frac{\lambda}{16\sigma^2}(\lambda^2+16t^2) & \frac{1}{32}\frac{\left(\lambda^2+16t^2\right)^2}{\sigma^4} & \frac{t}{\sigma^2}\left(\lambda^2+16t^2\right) \\ \frac{1}{\sigma_i^2} v_0^i & \frac{t}{\sigma_i^2} v_0^i & 2\lambda t n & \frac{t}{\sigma^2}\left(\lambda^2+16t^2\right) & \Sigma_{i=1}^n \frac{|v_0^i|^2}{\sigma_i^2} + 32nt^2 \end{pmatrix}.
$$

Remark 1.3.8. *Technically, the only possible case inside the* Maximal Strichartz Family of Gaussian Distributions *is when $R^T R = Id_{n\times n}$, since $R \in SO(n)$ (the spherical case, with $\sigma^2 = 1$). The form of the* Fisher Information Matrix*, in that case, simplifies to a lower dimension. Nevertheless, the computation performed in the way we did, gives the possibility to compute a distance (in some sense centred at the* Maximal Strichartz Family of Gaussian Distributions*) between members of the* Maximal Strichartz Family of Gaussian Distributions *and other Gaussian Distributions, for which the orthogonal matrix condition $R^T R = Id_{n\times n}$ is not necessarily satisfied. In particular, it can distinguish between Gaussians evolving through the PDE flow (See Section 1.3.2) and Gaussians which do not, because not correctly spread out in every spatial dimensions.*

As we said, Strichartz estimates are a way to measure the dispersion caused by the flow of the PDE to which they are related. In statistics, dispersion explains how stretched or squeezed a distribution is. A measure of statistical dispersion is a non-negative real number which is small for data which are very concentrated and increases as the data become more spread out. Common examples of measures of statistical dispersion are the variance, the standard deviation, the range and many others. Here, we connect the two closely related concepts (dispersion in statistics and PDEs) by introducing some measures of statistical dispersion like the *Index of Dispersion* in Definition 1.3.32 (see Section 1.3.4) which reflect the Dispersive PDE-nature of the *Maximal Strichartz Family of Gaussian Distributions*.

Definition 1.3.9. *Consider the norms* $||\cdot||_a$ *and* $||\cdot||_b$ *on the space of Variance-Covariance Matrices* Σ *and* $||\cdot||_c$ *on the space of mean values* μ. *We define the following* Index of Dispersion*:*

$$\mathcal{I}_M^{abc} := ||\Sigma(0)||_a \times \frac{||\Sigma(t)||_b}{||\mu(t)||_c^4} \tag{1.27}$$

with $t \neq 0$ *and where* $\mu(t)$

$$\mu(t) := x_0 + vt$$

while $\Sigma(t)$ *is given by*

$$\Sigma(t) := \frac{1}{4}\left(\lambda^2 + 16t^2\right) R^T R.$$

We call $\mathcal{I}_M^{abc}$ *the* abc-Dispersion Index of the Maximal Family of Gaussians *and we call*

$$\mathcal{I}_M^{a} := ||\Sigma(0)||_a$$

a-Static Dispersion Index of the Maximal Family of Gaussians*.*

We compute this *Index of Dispersion* for our family of distributions and show that it is consistent with PDE results. We refer to Definition 1.3.32 for more details.

Another important concept in probability and statistics is the one of *Stochastic Order*. A *Stochastic Order* is a way to consistently put a set of random variables in a sequence. Most of the *Stochastic Orders* are partial orders, in the sense that an order between the random variables exists, but not all the random variables can be put in the same sequence. Many different *Stochastic Orders* exist and have different applications. For more details on *Stochastic Orders*, we refer to [76]. Here, we use our *Index of Dispersion* to define a *Stochastic Order* on the *Maximal Strichartz Family of Gaussian Distributions* and see how there are natural ways of partially ranking the distributions of the family (See Section 1.3.5), in agreement with the flow of the PDE.

Definition 1.3.10. *Consider two random variables* X_1 *and* X_2 *such that* $\mu_{X_1}(\theta_1) = \mu_{X_2}(\theta_2)$, *for any* θ_1 *and* θ_2. *We say that the two random variables are* Ordered accordingly to their Dispersion Index $\mathcal{I}$ *if and only if the following condition is satisfied*

$$X_1 \prec X_2 \Leftrightarrow \mathcal{I}(X_2) \leq \mathcal{I}(X_1).$$

Remark 1.3.11. *In this definition the index $\mathcal{I}$ can vary accordingly to the context and the choices of the norms in the definition of the index.*

An important tool in our analysis is what we call $\frac{1}{\alpha}$-*Characteristic Function* (See Section 1.3.2) that we briefly anticipate from Chapter 2.

1.3.2 Construction of the *Maximal Strichartz Family of Gaussian Distributions*

This subsection is devoted to the construction of the *Maximal Strichartz Family of Gaussian Distributions.* The program is the following:

1. We define $\frac{1}{\alpha}$-*Characteristic Functions*;
2. We prove that if u_0 generates a probability distribution $p_0(x)$, then $u(t,x) = e^{it\Delta}u_0$ (See below for its precise definition) still defines a probability distribution $p_t(x) = |u(t,x)|^2$;
3. By means of $\frac{1}{\alpha}$-*Characteristic Functions*, we give the explicit expression of $u(t,x)$ the generator of the family;
4. We use symmetries and invariances to build the complete family $\mathcal{F}$.

1.3.2.1 The $\frac{1}{\alpha}$-*Characteristic Functions*

Following the program, we first need to introduce the tool of $\frac{1}{\alpha}$-*Characteristic Functions* to characterize $\mathcal{F}$. It is basically the Fourier Transform, but, differently from the *Characteristic Function*, the average is not taken with the pdf, but with a power of the pdf.

Definition 1.3.12. *Consider $u : \mathbb{R}^n \to \mathbb{C}$ to be a Schwartz function, namely a function belonging to the space*

$$\mathcal{S}(\mathbb{R}^n) := \{f \in C^\infty(\mathbb{R}^n) \mid \|f\|_{\alpha,\beta} < \infty \quad \forall \alpha, \beta\},$$

with α and β multi-indices, endowed with the following norm

$$\|f\|_{\alpha,\beta} = \sup_{x\in\mathbb{R}^n} \left| x^\alpha D^\beta f(x) \right|.$$

Moreover, suppose that

$$\int_{\mathbb{R}^n} |u|^\alpha = 1,$$

namely that $|u|^\alpha$ defines a continuous probability distribution function. Then, we define

$$\phi^u_\alpha(\xi) := \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} u(x)dx.$$

We call $\phi^u_\alpha(\xi)$ the $\frac{1}{\alpha}$-Characteristic Function *of u. Moreover, we define the* Inverse $\frac{1}{\alpha}$-Characteristic Function *by*

$$\psi^{\phi^u_\alpha}_\alpha(x) := \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \phi^u_\alpha(\xi)d\xi.$$

We refer to Chapter 2 for examples and properties of $\frac{1}{\alpha}$-*Characteristic Functions* and to the Chapter 2 applications of this tool. In particular, we notice that $\psi^{\phi^u_\alpha}_\alpha(x) = u(x)$.

Remark 1.3.13. *If u is essentially complex valued and for example $\alpha = n \in \mathbb{N}$, then there are n-distinct complex roots of $|u|^2$. In our discussion, this will not create to us any problem, because our process starts with u and produces $|u|^2$. We remark that the map $|u|^\alpha \mapsto u$ is a multivalued function. For this reason, we cannot reconstruct uniquely a generator, given the family that it generates. See formula (1.31) below for more details.*

Remark 1.3.14. *We could define* $\frac{1}{\alpha}$-Characteristic Functions *for more general functions $u : X \to F$ with X a locally compact Abelian group and F a general field. We do not pursue this direction here and we will leave it for a future work. We notice that $\phi^u_\alpha(\xi)$ can be considered also as a* $\frac{1}{\alpha}$-Expected Value

$$E^{\frac{1}{\alpha}}_{|u|^2}[e^{-ix\xi}] = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} u(x)dx.$$

1.3.2.2 Conservation of Mass and Flow on the space of probability measures

In this subsection, we show that if $p_0(x) = |u_0|^2$ defines a probability distribution, then also $p_t(x) = |e^{it\Delta}u_0|^2$ defines a probability distribution. This is mainly a consequence of the property of $e^{it\Delta}$ of being a unitary operator.

Theorem 1.3.15. *Consider $\mathcal{P}(\mathbb{R}^n)$, the set of all probability distributions on $\mathbb{R}^n$ and $u : (0, \infty) \times \mathbb{R}^n \to \mathbb{C}$ a solution to (1.9). Then u induces a flow in the space of probability distributions.*

Proof. Consider $u_0 : \mathbb{R}^n \to \mathbb{C}$ such that $||u||_{L^2(\mathbb{R}^n)} = 1$, so $p_0(x) := |u_0(x)|^2$ is a probability distribution on $\mathbb{R}^n$. Consider $u(t,x)$, the solution of (1.9) with initial datum u_0. Then

$$\partial_t \int_{\mathbb{R}^n} |u|^2 = \partial_t \int_{\mathbb{R}^n} u\bar{u} = \int_{\mathbb{R}^n} \partial_t(u\bar{u}) = \int_{\mathbb{R}^n} (\partial_t u\bar{u} + u\partial_t\bar{u}) \tag{1.28}$$

$$= \int_{\mathbb{R}^n} \Re\left[\bar{u}\left(\frac{i}{2}\Delta u\right) - u\left(\frac{i}{2}\Delta\bar{u}\right)\right] = 0. \tag{1.29}$$

So $\partial_t \int_{\mathbb{R}^n} |u|^2 = 0$ and hence

$$\int_{\mathbb{R}^n} |u(t,x)|^2 = \int_{\mathbb{R}^n} |u_0(x)|^2 = 1.$$

Therefore, for every $t \in \mathbb{R}$, $p(t,x) := |u(t,x)|^2$ is a probability distribution. □

$$\begin{array}{ccc} u_0 \in L^2(\mathbb{R}^n) & \xrightarrow{S(t)} & u(t) \in L^2(\mathbb{R}^n) \\ \Big\downarrow |\cdot|^2 & & \Big\downarrow |\cdot|^2 \\ p_0 \in \mathcal{P}(\mathbb{R}^n) & \xrightarrow[S(t)^*]{} & p(t) \in \mathcal{P}(\mathbb{R}^n) \end{array}$$

Remark 1.3.16. *This situation is in striking contrast with respect to the heat equation, where if you start with a probability distribution as initial datum, instantaneously the constraint of being a probability measure is broken.*

1.3.2.3 Fundamental Solution for The Linear Schrödinger Equation using $\frac{1}{\alpha}$-*Characteristic Functions*

Recall from Subsection 1.2.2.1 that the solution to the Linear Schrödinger Equation

$$i\partial_t u(t,x) = \Delta u(t,x), \quad (t,x) \in (0,\infty) \times \mathbb{R}^n, \tag{1.30}$$

with initial datum $u_0(x) = e^{-|x|^2} \in \mathcal{S}(\mathbb{R}^n)$ is given by

$$u(t,x) = (1-4it)^{-n/2} e^{-\frac{|x|^2}{1-4it}},$$

which produces the following probability density function

$$p(t,x) = \left(\frac{\pi}{2}\right)^{-\frac{n}{2}} |1+16t^2|^{-n/2} e^{-\frac{2|x|^2}{1+16t^2}} \tag{1.31}$$

which is going to be the generator of the family of distributions $\mathcal{F}$, due to Theorem 1.3.15.

1.3.2.4 Strichartz Estimates and their symmetries

Recall from Section 1.2.2.2 Theorem 1.2.10, namely that Strichartz Estimates are invariant by the following set of symmetries.

Lemma 1.3.17. *[53] Let $\mathcal{G}$ be the group of transformations generated by:*

- *space-time translations: $u(t,x) \mapsto u(t+t_0, x+x_0)$, with $t_0 \in \mathbb{R}$, $x_0 \in \mathbb{R}^n$;*
- *parabolic dilations: $u(t,x) \mapsto u(\lambda^2 t, \lambda x)$, with $\lambda > 0$;*
- *change of scale: $u(t,x) \mapsto \mu u(t,x)$, with $\mu > 0$;*
- *space rotations: $u(t,x) \mapsto u(t, Rx)$, with $R \in SO(n)$;*
- *phase shifts: $u(t,x) \mapsto e^{i\theta} u(t,x)$, with $\theta \in \mathbb{R}$;*
- *Galilean transformations:*

  $$u(t,x) \mapsto e^{\frac{i}{4}\left(|v|^2 t + 2v\cdot x\right)} u(t, x+tv),$$

  with $v \in \mathbb{R}^n$.

Then, if u solves equation (1.9) and $g \in \mathcal{G}$, also $v = g \circ u$ solves equation (1.9). Moreover, the constants $S_h(n,q,r)$, $S_d(n,q,r)$ and $S_i(n,q,r,\tilde{q},\tilde{r})$ are left unchanged by the action of $\mathcal{G}$.

Not all these symmetries leave invariant the set of probability distributions $\mathcal{P}(\mathbb{R}^n)$. Therefore, we need to reduce the set of symmetries in our treatment and, in particular, we need to combine the scaling and the parabolic dilations in order to have all the family inside the space of probability distributions $\mathcal{P}(\mathbb{R}^n)$.

Lemma 1.3.18. *Consider $u_{\mu,\lambda} = \mu u(\lambda^2 t, \lambda x)$ such that $u(t,x) \in \mathcal{P}(\mathbb{R}^n)$ maximizes (1.22), then $\mu = \lambda^{n/2}$.*

Proof.

$$1 = ||u_\lambda||^2_{L^2(\mathbb{R}^n)} = \mu^2 \int_{\mathbb{R}^n} |u(\lambda^2 t, \lambda x)|^2 dx = \mu^2 \lambda^{-n} ||u||^2_{L^2(\mathbb{R}^n)} = \mu^2 \lambda^{-n},$$

so $\mu = \lambda^{n/2}$. □

Remark 1.3.19. *We notice that some of the symmetries can be seen just at the level of the generator of the family u, but not by the family of probability distributions $p_t(x)$. For example the phase shifts $u(t,x) \mapsto e^{i\theta} u(t,x)$, with $\theta \in \mathbb{R}$ give rise to the same probability distribution function because $|e^{i\theta} u(t,x)|^2 = |u(t,x)|^2$ and, partially, the Galilean transformations $u(t,x) \mapsto e^{\frac{i}{4}\left(|v|^2 t + 2v\cdot x\right)} u(t, x+tv)$, with $v \in \mathbb{R}^n$ reduces to a space translation with $x_0 = vt$, since $\left| e^{\frac{i}{4}\left(|v|^2 t + 2v\cdot x\right)} u(t, x+tv) \right|^2 = |u(t, x+tv)|^2$. In some sense, the parameter θ can be seen as a* latent variable.

Therefore, we have the complete set of probability distributions induced by the generator $u(t,x)$.

Theorem 1.3.20. *Consider $p_t(x) = |u(t,x)|^2$ a probability distribution function generated by $u(t,x)$ (see Subsection 1.2.2.1). Let $\mathcal{S}$ be the group of transformations generated by:*

- *inertial-space translations and time translations: $p(t,x) \mapsto p(t+t_0, x+x_0+vt)$, with $t_0 \in \mathbb{R}$, $x_0 \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$;*
- *scaling-parabolic dilations: $u(t,x) \mapsto \lambda^n u(\lambda^2 t, \lambda x)$, with $\lambda > 0$;*
- *space rotations: $u(t,x) \mapsto u(t, Rx)$, with $R \in SO(n)$;*

Then, if u solves equation (1.9) and $g \in \mathcal{S}$, also $v = g \circ u$ solves equation (1.9), $q_t(x) = |v(t,x)|^2$ is still a probability distribution for every $g \in \mathcal{S}$ and the constant $S_h(n,q,r)$ is left unchanged by the action of $\mathcal{S}$.

This theorem produces the following definition:

Definition 1.3.21. *We call* Maximal Strichartz Family of Gaussian Distributions *the following family of distributions:*

$$\mathcal{F} := \Big\{ p_t(x) = \left(\frac{\pi}{2}\right)^{-\frac{n}{2}} |R^T R|^{-\frac{1}{2}} |\lambda^2 + 16t^2|^{-n/2} e^{-\frac{2(x-x_0-vt)^T (R^T R)^{-1} (x-x_0-vt)}{\lambda^2+16t^2}} \tag{1.32}$$

$$(t,\lambda) \in \mathbb{R}^2, (x_0, v) \in \mathbb{R}^n \times \mathbb{R}^n, R \in SO(n) \Big\}. \tag{1.33}$$

Remark 1.3.22. *Let $p(t,x)$ be the pdf defined in equation (1.31). Then, choose $\tilde{p}_t(x) \in \mathcal{F}$ with $R = Id$, $x_0 = v_0 = 0$ and $\lambda = 0$. This implies: $\tilde{p}_t(x) = p(t,x) \in \mathcal{F}$. For this reason, we call $p(t,x)$ the* Family Generator *of $\mathcal{F}$. We notice also that, in the definition of the family and with respect to Theorem 1.3.20, we used as scale parameter $\lambda^{1/2}$ instead of λ. This is done without loss of generality, since $\lambda > 0$.*

Right away we can compute the *Variance-Covariance Matrix* and *Mean Vector* of the family.

Corollary 1.3.23. *Suppose $\mathbf{X}$ is a random variable with pdf $p_t(x) \in \mathcal{F}$. Then its* Expected Value *is*

$$E[\mathbf{X}] := \mu = x_0 + vt$$

and its Variance *is*

$$\Sigma = \frac{1}{4}(\lambda^2 + 16t^2)(R^T R).$$

Proof. The proof is a direct computation. □

Remark 1.3.24. *We see here that, differently from the general family of Gaussian distributions, here the* Mean Vector *and the* Variance-Covariance Matrix *are related by a parameter, which represents the time flow.*

1.3.3 The Fisher Information Metric of the Maximal Strichartz Family $\mathcal{F}$

Information geometry is a branch of mathematics that applies the techniques of differential geometry to the field of statistics and probability theory. This is done by interpreting probability distributions of a statistical model as the points of a Riemannian manifold, forming in this way a statistical manifold. The Fisher information metric provides a natural Riemannian metric for this manifold, but it is not the only possible one. With this tool, we can define and compute meaningful distances between probability distributions, both in the discrete and continuous case. Crucial is then the set of parameters on which a certain family of distributions is indexed and the geometrical structure of the parameter set. We refer to [1] for a general reference on information geometry. The first one to introduce the notion of distance between two probability distributions has been Rao in [98] (see also [4]), who used the Fisher Information Matrix as a Riemannian Metric on the space of parameters.

In this subsection, we restrict our attention to the Fisher Information Metric of the *Maximal Strichartz Family of Gaussian Distributions* $\mathcal{F}$ and provide details on the additional structure that the family has with respect to the hyperbolic model of the general Family of Gaussian Distributions. See for example [34], [35] and [77].

1.3.3.1 The Fisher Information Metric for the Multivariate Gaussian Distribution

First, we give the general definition of the Fisher Information Metric:

Definition 1.3.25. *Consider a statistical manifold $\mathcal{S}$, with coordinates given by $\theta = (\theta_1, \theta_2, \cdots, \theta_n)$ and with probability density function $p(x;\theta)$. Here, x is a specific observation of the discrete or continuous random variables X. The probability is normalized, so that $\int_X p(x,\theta)dx = 1$ for every $\theta \in \mathcal{S}$. The* Fisher Information Metric *I_{ij} is defined by the following formula:*

$$I_{ij}(\theta) = \int_X \frac{\partial \log p(x,\theta)}{\partial \theta_i} \frac{\partial \log p(x,\theta)}{\partial \theta_j} p(x,\theta)\, dx. \tag{1.34}$$

Remark 1.3.26. *The integral is performed over all values x that the random variable X can take. Again, the variable θ is understood as a coordinate on the statistical manifold $\mathcal{S}$, intended as a* Riemannian Manifold*. Under certain regularity conditions (any that allows integration by parts), I_{ij} can be rewritten as*

$$I_{ij}(\theta) = -\int_X \frac{\partial^2 logp(x,\theta)}{\partial \theta_i \partial \theta_j} p(x,\theta)\, dx = -\mathrm{E}\left[\frac{\partial^2 logp(x,\theta)}{\partial \theta_i \partial \theta_j}\right]. \tag{1.35}$$

Now, to compute explicitly the *Fisher Information Matrix* of the family $\mathcal{F}$, we use the following theorem that you can find in [86].

Theorem 1.3.27. *The* Fisher Information Matrix *for an n-variate Gaussian distribution can be computed in the following way. Let*

$$\mu(\theta) = \left[\mu_1(\theta), \mu_2(\theta), \ldots, \mu_N(\theta)\right]^{\mathrm{T}}$$

be the vector of Expected Values *and let $\Sigma(\theta)$ be the* Variance-Covariance Matrix*. Then, the typical element $\mathcal{I}_{i,j}, 0 \le i, j < n$, of the* Fisher Information Matrix *for a random variable $X \sim \mathcal{N}(\mu(\theta), \Sigma(\theta))$ is:*

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} + \frac{1}{2} \operatorname{tr}\left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j}\right),$$

where $(..)^{\mathrm{T}}$ *denotes the transpose of a vector,* $tr(..)$ *denotes the trace of a square matrix, and:*

$$\frac{\partial\mu}{\partial\theta_i} = \begin{bmatrix} \frac{\partial\mu_1}{\partial\theta_i} & \frac{\partial\mu_2}{\partial\theta_i} & \cdots & \frac{\partial\mu_n}{\partial\theta_i} \end{bmatrix}^{\mathrm{T}};$$

and

$$\frac{\partial\Sigma}{\partial\theta_i} = \begin{bmatrix} \frac{\partial\Sigma_{1,1}}{\partial\theta_i} & \frac{\partial\Sigma_{1,2}}{\partial\theta_i} & \cdots & \frac{\partial\Sigma_{1,n}}{\partial\theta_i} \\ \frac{\partial\Sigma_{2,1}}{\partial\theta_i} & \frac{\partial\Sigma_{2,2}}{\partial\theta_i} & \cdots & \frac{\partial\Sigma_{2,n}}{\partial\theta_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\Sigma_{n,1}}{\partial\theta_i} & \frac{\partial\Sigma_{n,2}}{\partial\theta_i} & \cdots & \frac{\partial\Sigma_{n,n}}{\partial\theta_i} \end{bmatrix}.$$

Now, we have just to compute the *Fisher Information Matrix* entry by entry, following the theorem. We recall here that we are considering the following family of Gaussian Distributions:

$$\mathcal{F} := \Big\{ p(t,x) = \left(\frac{\pi}{2}\right)^{-\frac{n}{2}} |R^T R|^{-\frac{1}{2}} |\lambda^2 + 16t^2|^{-n/2} e^{-\frac{2(x-x_0-vt)^T (R^T R)^{-1}(x-x_0-vt)}{\lambda^2+16t^2}} :$$
$$(t,\lambda) \in \mathbb{R}^2, (x_0, v) \in \mathbb{R}^n \times \mathbb{R}^n, R \in SO(n) \Big\}.$$

and that, in particular, we have that the *Expected Value* of a random variable X with distribution belonging to the family $\mathcal{F}$ is given by

$$\mu := x_0 + vt,$$

while the *Variance-Covariance Matrix* is given by

$$\Sigma := \frac{1}{4}\left(\lambda^2 + 16t^2\right) R^T R.$$

Remark 1.3.28. *We remark again that* μ *and* Σ *depend on some common parameters, like the time* t*.*

1.3.3.2 Proof of Theorem 1.3.7: The Spherical Multivariate Gaussian Distribution

Here, we consider the case in which $R^T R = \sigma^2 Id_{n\times n}$, namely the case where the *Variance-Covariance Matrix* is given by $\Sigma := \frac{1}{4}\left(\lambda^2 + 16t^2\right)\sigma^2 Id_{n\times n}$. In this case, the vector of parameters θ is given by

$$\theta := (x_0^T, v_0^T, \lambda, \sigma^2, t)^T,$$

with x_0 and v_0 are $n \times 1$, while λ, σ^2, t are scalars. In order to fix the notation, we call $(\theta_1, \ldots, \theta_n)^T := (x_0, \ldots, x_n)$, $(\theta_{n+1}, \ldots, \theta_{2n})^T := (v_0^0, \ldots, v_0^n)$, $\theta_{2n+1} = \lambda$, $\theta_{2n+2} = \sigma^2$ and $\theta_{2n+3} = t$. Now, we want to compute all the coefficients of I_{ij}. We use the symmetry of the information matrix I_{ij}, so $I_{ij} = I_{ji}$. The relevant coefficients are the following:

- $i = 1, \ldots, n$, $j = 1, \ldots, n$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} = (0, \ldots, 1^i, \ldots, 0)^T \frac{1}{\sigma^2} Id_{n \times n} (0, \ldots, 1^j, \ldots, 0) = \frac{1}{\sigma^2} \delta_{ij};$$

- $i = 1, \ldots, n$, $j = n+1, \ldots, 2n$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} = (0, \ldots, 1^i, \ldots, 0)^T \frac{1}{\sigma^2} Id_{n \times n} (0, \ldots, t^j, \ldots, 0) = \frac{t}{\sigma^2} \delta_{ij};$$

- $i = 1, \ldots, n$, $j = 2n+1$

$$\mathcal{I}_{i,j} = 0,$$

  because μ does not depend on λ and Σ does not depend on x_0;

- $i = 1, \ldots, n$, $j = 2n+2$

$$\mathcal{I}_{i,j} = 0,$$

  because μ does not depend on σ^2 and Σ does not depend on x_0;

- $i = 1, \ldots, n$, $j = 2n+3$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} = (0, \ldots, 1^i, \ldots, 0)^T \frac{1}{\sigma^2} Id_{n \times n} (v_0^0, \ldots, v_0^n) = \frac{1}{\sigma^2} v_0^i;$$

- $i = n+1, \ldots, 2n$, $j = n+1, \ldots, 2n$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} = (0, \ldots, t^i, \ldots, 0)^T \frac{1}{\sigma^2} Id_{n \times n} (0, \ldots, t^j, \ldots, 0) = \frac{t^2}{\sigma^2} \delta_{ij};$$

- $i = n+1, \ldots, 2n$, $j = 2n+1$

$$\mathcal{I}_{i,j} = 0,$$

because μ does not on λ and Σ does not depend on v_0;

- $i = n+1, \dots, 2n,\ j = 2n+2$

$$\mathcal{I}_{i,j} = 0,$$

because μ does not depend on σ^2 and Σ does not depend on v_0 ;

- $i = n+1, \dots, 2n,\ j = 2n+3$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i}\Sigma^{-1}\frac{\partial \mu}{\partial \theta_j} = (0, \dots, 1^i, \dots, 0)^T \frac{1}{\sigma^2} Id_{n\times n}(v_0^0, \dots, v_0^n) = \frac{t}{\sigma^2} v_0^i;$$

- $i = 2n+1,\ j = 2n+1$

$$\mathcal{I}_{i,j} = \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_j}\right) = \frac{1}{2} tr\left(\frac{1}{\sigma^2} Id \frac{2\lambda\sigma^2}{4} Id \frac{1}{\sigma^2} Id \frac{2\lambda\sigma^2}{4} Id\right)$$
$$= \frac{\lambda^2}{8} tr(Id) = \frac{\lambda^2 n}{8};$$

- $i = 2n+1,\ j = 2n+2$

$$\begin{aligned}\mathcal{I}_{i,j} = \quad & \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_j}\right) = \frac{1}{2} tr\left(\frac{1}{\sigma^2} Id \frac{2\lambda\sigma^2}{4} Id \frac{1}{\sigma^2} Id \frac{\lambda^2 + 16t^2}{4} Id\right) \\ & = n\frac{\lambda}{16\sigma^2}(\lambda^2 + 16t^2);\end{aligned}$$

- $i = 2n+1,\ j = 2n+3$

$$\mathcal{I}_{i,j} = \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_j}\right) = \frac{1}{2} tr\left(\frac{1}{\sigma^2} Id \frac{2\lambda\sigma^2}{4} Id \frac{1}{\sigma^2} Id \frac{8t\sigma^2}{4} Id\right) = 2\lambda t n;$$

- $i = 2n+2,\ j = 2n+2$

$$\begin{aligned}\mathcal{I}_{i,j} = \quad & \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_j}\right) = \frac{1}{2} tr\left(\frac{1}{\sigma^2} Id \frac{\lambda^2 + 16t^2}{4} Id \frac{1}{\sigma^2} Id \frac{\lambda^2 + 16t^2}{4} Id\right) \\ & = \frac{n}{32}\frac{(\lambda^2 + 16t^2)^2}{\sigma^4};\end{aligned}$$

- $i = 2n+2,\ j = 2n+3$

$$\mathcal{I}_{i,j} = \frac{1}{2}\,\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right) = \frac{1}{2}tr\left(\frac{1}{\sigma^2}Id\frac{\lambda^2+16t^2}{4}Id\frac{1}{\sigma^2}Id8t\sigma^2Id\right)$$
$$= \frac{nt}{\sigma^2}\Big(\lambda^2+16t^2\Big);$$

- $i = 2n+3,\ j = 2n+3$

$$\begin{aligned}\mathcal{I}_{i,j} = \quad & \frac{\partial\mu^{\mathrm{T}}}{\partial\theta_i}\Sigma^{-1}\frac{\partial\mu}{\partial\theta_j} + \frac{1}{2}\,\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right) = \\ & (v_0^0,\ldots,v_0^n)^T\frac{1}{\sigma^2}Id_{n\times n}(v_0^0,\ldots,v_0^n) + \frac{1}{2}tr\left(\frac{1}{\sigma^2}Id8t\sigma^2Id\frac{1}{\sigma^2}Id8t\sigma^2Id\right) = \\ & = \frac{|v_0|^2}{\sigma^2} + 32nt^2.\end{aligned}$$

In conclusion, we have

$$I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}Id & \frac{t}{\sigma^2}Id & 0 & 0 & \frac{1}{\sigma^2}v_0 \\ \frac{t}{\sigma^2}Id & \frac{t^2}{\sigma^2}Id & 0 & 0 & \frac{t}{\sigma^2}v_0 \\ 0 & 0 & \frac{\lambda^2 n}{8} & n\frac{\lambda}{16\sigma^2}(\lambda^2+16t^2) & 2\lambda tn \\ 0 & 0 & n\frac{\lambda}{16\sigma^2}(\lambda^2+16t^2) & \frac{n}{32}\frac{\Big(\lambda^2+16t^2\Big)^2}{\sigma^4} & \frac{nt}{\sigma^2}\Big(\lambda^2+16t^2\Big) \\ \frac{1}{\sigma^2}v_0 & \frac{t}{\sigma^2}v_0 & 2\lambda tn & \frac{nt}{\sigma^2}\Big(\lambda^2+16t^2\Big) & \frac{|v_0|^2}{\sigma^2}+32nt^2 \end{pmatrix}.$$

1.3.3.3 Proof of Theorem 1.3.7: The Elliptical Multivariate Gaussian Distribution

We define

$$\frac{1}{\sigma_i^2} Id := \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix}.$$

We define also

$$\frac{1}{\sigma^2} := \Sigma_{i=1}^n \frac{1}{\sigma_i^2}$$

and

$$\frac{1}{\sigma^4} := \Sigma_{i=1}^n \frac{1}{\sigma_i^4}.$$

Using this notations, we are going to compute the matrix I_{ij}. The relevant coefficients are the following:

- $i = 1, \ldots, n$, $j = 1, \ldots, n$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} = (0, \ldots, 1^i, \ldots, 0)^T \frac{1}{\sigma_i^2} Id_{n \times n} (0, \ldots, 1^j, \ldots, 0) = \frac{1}{\sigma_i^2} \delta_{ij};$$

- $i = 1, \ldots, n$, $j = n+1, \ldots, 2n$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} = (0, \ldots, 1^i, \ldots, 0)^T \frac{1}{\sigma_i^2} Id_{n \times n} (0, \ldots, t^j, \ldots, 0) = \frac{t}{\sigma_i^2} \delta_{ij};$$

- $i = 1, \ldots, n$, $j = 2n+1$

$$\mathcal{I}_{i,j} = 0,$$

  because μ does not depend on λ and Σ does not depend on x_0;

- $i = 1, \ldots, n$, $j = 2n+2$

$$\mathcal{I}_{i,j} = 0,$$

because μ does not depend on σ^2 and Σ does not depend on x_0;

- $i = 1, \ldots, n$, $j = 2n+3$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} = (0, \ldots, 1^i, \ldots, 0)^T \frac{1}{\sigma^2} Id_{n \times n} (v_0^0, \ldots, v_0^n) = \frac{1}{\sigma_i^2} v_0^i$$
$$= (v_0^0/\sigma_1^2, \ldots, v_0^n/\sigma_n^2);$$

- $i = n+1, \ldots, 2n$, $j = n+1, \ldots, 2n$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} = (0, \ldots, t^i, \ldots, 0)^T \frac{1}{\sigma^2} Id_{n \times n} (0, \ldots, t^j, \ldots, 0) = \frac{t^2}{\sigma_i^2} \delta_{ij};$$

- $i = n+1, \ldots, 2n$, $j = 2n+1$

$$\mathcal{I}_{i,j} = 0,$$

because μ does not depend on λ and Σ does not depend on v_0;

- $i = n+1, \ldots, 2n$, $j = 2n+2$

$$\mathcal{I}_{i,j} = 0,$$

because μ does not depend on σ^2 and Σ does not depend on v_0;

- $i = n+1, \ldots, 2n$, $j = 2n+3$

$$\mathcal{I}_{i,j} = \frac{\partial \mu^{\mathrm{T}}}{\partial \theta_i} \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} = (0, \ldots, 1^i, \ldots, 0)^T \frac{1}{\sigma^2} Id_{n \times n} (v_0^0, \ldots, v_0^n) = \frac{t}{\sigma_i^2} v_0^i$$
$$= t(v_0^0/\sigma_1^2, \ldots, v_0^n/\sigma_n^2);$$

- $i = 2n+1$, $j = 2n+1$

$$\mathcal{I}_{i,j} = \frac{1}{2} \operatorname{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \right) = \frac{1}{2} tr \left(\frac{1}{\sigma^2} Id \frac{2\lambda\sigma^2}{4} Id \frac{1}{\sigma^2} Id \frac{2\lambda\sigma^2}{4} Id \right)$$
$$= \frac{\lambda^2}{8} tr(Id) = \frac{\lambda^2 n}{8};$$

- $i = 2n+1$, $j = 2n+2$

$$
\begin{aligned}
\mathcal{I}_{i,j} &= \frac{1}{2}\,\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right) = \frac{1}{2}tr\left(\frac{1}{\sigma_i^2}Id\frac{2\lambda\sigma_i^2}{4}Id\frac{1}{\sigma_i^2}Id\frac{\lambda^2+16t^2}{4}Id\right)\\
&= \Sigma_{i=1}^n\frac{\lambda}{16\sigma_i^2}(\lambda^2+16t^2) = \frac{\lambda}{16\sigma^2}(\lambda^2+16t^2);
\end{aligned}
$$

- $i = 2n+1,\ j = 2n+3$

$$
\begin{aligned}
\mathcal{I}_{i,j} = \quad & \frac{1}{2}\,\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right)\\
&= \frac{1}{2}tr\left(\frac{1}{\sigma_i^2}Id\frac{2\lambda\sigma_i^2}{4}Id\frac{1}{\sigma_i^2}Id\frac{8t\sigma_i^2}{4}Id\right) = 2\lambda tn;
\end{aligned}
$$

- $i = 2n+2,\ j = 2n+2$

$$
\begin{aligned}
\mathcal{I}_{i,j} = \quad & \frac{1}{2}\,\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right) = \frac{1}{2}tr\left(\frac{1}{\sigma_i^2}Id\frac{\lambda^2+16t^2}{4}Id\frac{1}{\sigma_i^2}Id\frac{\lambda^2+16t^2}{4}Id\right)\\
&= \Sigma_{i=1}^n\frac{1}{32}\frac{\left(\frac{\lambda^2+16t^2}{4}\right)^2}{\sigma_i^4} = \frac{1}{32}\frac{\left(\frac{\lambda^2+16t^2}{4}\right)^2}{\sigma^4};
\end{aligned}
$$

- $i = 2n+2,\ j = 2n+3$

$$
\begin{aligned}
\mathcal{I}_{i,j} = \quad & \frac{1}{2}\,\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right) = \frac{1}{2}tr\left(\frac{1}{\sigma_i^2}Id\frac{\lambda^2+16t^2}{4}Id\frac{1}{\sigma_i^2}Id8t\sigma^2Id\right)\\
&= \Sigma_{i=1}^n\frac{t}{\sigma_i^2}\left(\lambda^2+16t^2\right) = \frac{t}{\sigma^2}\left(\lambda^2+16t^2\right);
\end{aligned}
$$

- $i = 2n+3,\ j = 2n+3$

$$
\begin{aligned}
\mathcal{I}_{i,j} = \quad & \frac{\partial\mu^{\mathrm{T}}}{\partial\theta_i}\Sigma^{-1}\frac{\partial\mu}{\partial\theta_j} + \frac{1}{2}\,\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right) =\\
& (v_0^0,\ldots,v_0^n)^T\frac{1}{\sigma_i^2}Id_{n\times n}(v_0^0,\ldots,v_0^n) + \frac{1}{2}tr\left(\frac{1}{\sigma_i^2}Id8t\sigma_i^2Id\frac{1}{\sigma_i^2}Id8t\sigma_i^2Id\right) =\\
& = \Sigma_{i=1}^n|v_0^i|^2/\sigma_i^2 + 32nt^2.
\end{aligned}
$$

In conclusion, we have

$$
I(\theta) = \begin{pmatrix}
\frac{1}{\sigma_i^2} Id & \frac{t}{\sigma_i^2} Id & 0 & 0 & \frac{1}{\sigma_i^2} v_0^i \\
\frac{t}{\sigma_i^2} Id & \frac{t^2}{\sigma_i^2} Id & 0 & 0 & \frac{t}{\sigma_i^2} v_0^i \\
0 & 0 & \frac{\lambda^2 n}{8} & \frac{\lambda}{16\sigma^2}(\lambda^2 + 16t^2) & 2\lambda t n \\
0 & 0 & \frac{\lambda}{16\sigma^2}(\lambda^2 + 16t^2) & \frac{1}{32}\frac{\left(\lambda^2+16t^2\right)^2}{\sigma^4} & \frac{t}{\sigma^2}\left(\lambda^2 + 16t^2\right) \\
\frac{1}{\sigma_i^2} v_0^i & \frac{t}{\sigma_i^2} v_0^i & 2\lambda t n & \frac{t}{\sigma^2}\left(\lambda^2 + 16t^2\right) & \Sigma_{i=1}^n |v_0^i|^2/\sigma_i^2 + 32nt^2
\end{pmatrix}.
$$

This concludes the proof of Theorem 1.3.7.

1.3.3.4 The General Multivariate Gaussian Distribution

As pointed out in [34] and [35], for general multivariate normal distributions, the explicit form of the Fisher distance has not been computed in closed form yet even in the simple case where the parameters are $t = 0, \lambda = 0, v_0 = 0$. From a technical point of view, as pointed out in [34] and [35], the main difficulty arises from the fact that the sectional curvatures of the Riemannian manifold induced by $\mathcal{F}$ and endowed with the *Fisher Information Metric*, are not all constant.

We remark again here that the distance induced by our *Fisher Information Matrix* is centred at the *Maximal Strichartz Family of Gaussian Distributions*, to enlighten the difference between members of the *Maximal Strichartz Family of Gaussian Distributions* and other Gaussian Distributions, for which $R^T R = Id_{n\times n}$ is not necessarily satisfied. In particular, our metric distinguishes between Gaussians evolving through the PDE flow (See Section 1.3.2) and Gaussians who do not, because not correctly spread out in every spatial dimensions.

Remark 1.3.29. *We say that two parameters α and β are orthogonal if the elements of the corresponding rows and columns of the* Fisher Information Matrix *are zero. Orthogonal parameters are easy to deal with in the sense that their maximum likelihood estimates are independent and can be calculated separately. In particular, for our family $\mathcal{F}$ the parameters x_0 and v_0 are both orthogonal to both the parameters λ and σ^2. Some partial results, for example when either mean or variance are kept constant can be deduced. See for example [34], [35] and [77].*

Remark 1.3.30. *The* Fisher Information Metric *is not the only possible choice to compute distances between pdfs of the family of* Gaussian Distributions*. For example in [77], the* *authors parametrize the family of normal distribution as the symmetric space $SL(n+1)/SO(n+1)$ endowed with the following metric*

$$ds^2 = tr(\Sigma^{-1}d\Sigma\Sigma^{-1}d\Sigma) - \frac{1}{n+1}(tr(\Sigma^{-1}d\Sigma))^2 + \frac{1}{2}d\mu^T\Sigma^{-1}d\mu.$$

Moreover, the authors in [77] computed the Riemann Curvature Tensor *of the metric and, in any dimension, the distance between two normal distributions with the same mean and different variance and also the distance between two normal distributions with the same variance and different mean.*

Remark 1.3.31. *If we consider just the submanifold given by the restriction to the coordinates $i = 1, \dots, n$ and $i = 2n+2$ on the ellipse $\lambda^2 + 16t^2 = 4$ we recover the hyperbolic distance*

$$ds^2_{\mathbf{H}} := \frac{d\mu^2}{\sigma^2} + \frac{d\sigma^2}{2\sigma^4}.$$

The geometry however does not seem the one of a product space, at least considering the fact that mixed entries are not zero, in our parametrization.

1.3.4 Overdispersion, Equidispersion and UnderDispersion for the Family $\mathcal{F}$

As we said, Strichartz estimates are a way to measure the dispersion caused by the flow of the PDE to which they are related. In statistics, dispersion explains how spread out a distribution is. In this subsection, we connect the two closely related concepts (dispersion in statistics and PDEs) by introducing some measures of statistical dispersion like the *Index of Dispersion* in Definition 1.3.32 (see Section 1.3.4) which reflect the Dispersive PDE-nature of the *Maximal Strichartz Family of Gaussian Distributions.* We compute this *Index of Dispersion* for our family of distributions and show that it is consistent with PDE results.

Definition 1.3.32. *Consider the norms $||\cdot||_a$ and $||\cdot||_b$ on the space of Variance-Covariance Matrices Σ and $||\cdot||_c$ on the space of mean values μ. We define the following* Index of Dispersion*:*

$$\mathcal{I}_M^{abc} := ||\Sigma(0)||_a \times \frac{||\Sigma(t)||_b}{||\mu(t)||_c^4} \tag{1.36}$$

with $t \neq 0$ and where $\mu(t)$

$$\mu(t) := x_0 + vt$$

while $\Sigma(t)$ *is given by*

$$\Sigma(t) := \frac{1}{4}\left(\lambda^2 + 16t^2\right) R^T R.$$

We call $\mathcal{I}_M^{abc}$ *the* abc-Dispersion Index of the Maximal Family of Gaussians *and we call*

$$\mathcal{I}_M^a := ||\Sigma(0)||_a$$

a-Static Dispersion Index of the Maximal Family of Gaussians*. Moreover, we say that the distribution is:*

- abc-overdispersed*, if* $\mathcal{I}_M^{abc} < 1$*;*
- abc-equidispersed*, if* $\mathcal{I}_M^{abc} = 1$*;*
- abc-underdispersed*, if* $\mathcal{I}_M^{abc} > 1$*.*

Analogously, we say that the distribution is:

- a-overdispersed*, if* $\mathcal{I}_M^{a} < 1$*;*
- a-equidispersed*, if* $\mathcal{I}_M^{a} = 1$*;*
- a-underdispersed*, if* $\mathcal{I}_M^{a} > 1$*.*

Here, we discuss some particular cases and compute the dispersion indexes $\mathcal{I}_M^{abc}$ and $\mathcal{I}_M^a$ for certain specific norms $||\cdot||_a$, $||\cdot||_b$ and $||\cdot||_c$.

- In the case $t = 0$, the *a-Static Dispersion Index of the Maximal Family of Gaussians* that we choose, is given by the variance of the distribution. We choose $||\Sigma||_a = det(\Sigma)$ and so we get
$$\mathcal{I}_M^a = ||\Sigma||_a = det(\Sigma) = \left(\frac{1}{4}\lambda^2\right)^n det(R^T R).$$
Now, in the spherical case $R^T R = \sigma^2 Id$, one gets
$$\mathcal{I}_M^a = \left(\frac{1}{4}\lambda^2\sigma^2\right)^n.$$
So, the distribution is
  - *a-overdispersed* if $\frac{1}{4}\lambda^2\sigma^2 < 1$;

– *a-equidispersed* if $\frac{1}{4}\lambda^2\sigma^2 = 1$;

– *a-underdispersed* if $\frac{1}{4}\lambda^2\sigma^2 > 1$.

Therefore, with $||\Sigma||_a = det(\Sigma)$ the type of dispersion does not depend on the dimension n.

Remark 1.3.33. *In the strictly Strichartz case $\sigma^2 = 1$, we have that the dispersion is measured just by the scaling factor λ.*

Choosing instead $||\Sigma||_a = tr(\Sigma)$ as *a-Static Dispersion Index of the Maximal Family of Gaussians*, we have some small differences:

$$\mathcal{I}_M^a = ||\Sigma||_a = tr(\Sigma) = \left(\frac{1}{4}\lambda^2\right) tr(R^T R).$$

Now, in the spherical case $R^T R = \sigma^2 Id$, we get

$$\mathcal{I}_M^a = n\left(\frac{1}{4}\lambda^2\sigma^2\right).$$

So, the distribution is

– *a-overdispersed* if $n\frac{1}{4}\lambda^2\sigma^2 < 1$;

– *a-equidispersed* if $n\frac{1}{4}\lambda^2\sigma^2 = 1$;

– *a-underdispersed* if $n\frac{1}{4}\lambda^2\sigma^2 > 1$.

So with $||\Sigma||_a = tr(\Sigma)$ the type of dispersion does depend on the dimension n.

- In the case $t \in \mathbb{R}$, when t is different from zero, we can express Σ as a function of μ. In fact we have:

$$t = \frac{(\mu - x_0)\cdot v_0}{|v_0|^2}$$

and so

$$\Sigma(\mu) = \frac{1}{4}\left(\lambda^2 + 16\Big(\frac{(\mu - x_0)\cdot v_0}{|v_0|^2}\Big)^2\right) R^T R.$$

For example, if now we choose $x_0 = 0$ and $\mu = v_0 = (1, 0, \dots, 0)^T$ we get

$$\Sigma(\mu) = \frac{1}{4}\Big(\lambda^2 + 16\Big) R^T R,$$

so for $||\Sigma||_a = ||\Sigma||_b = det(\Sigma)$ and $||\mu||_c^2 = \Sigma_{1=1}^n |\mu_i|^2$ we get, in the spherical case $R^T R = \sigma^2 Id$,

$$\mathcal{I}_M^{abc} = \left(\frac{1}{4}\lambda^2\sigma^2\right)^n \times \left(\frac{1}{4}(\lambda^2+16)\sigma^2\right)^n.$$

So the distribution is

- *abc-overdispersed* if $\left(\frac{1}{4}\lambda^2\sigma^2\right)^n \times \left(\frac{1}{4}(\lambda^2+16)\sigma^2\right)^n < 1$;
- *abc-equidispersed* if $\left(\frac{1}{4}\lambda^2\sigma^2\right)^n \times \left(\frac{1}{4}(\lambda^2+16)\sigma^2\right)^n = 1$;
- *abc-underdispersed* if $\left(\frac{1}{4}\lambda^2\sigma^2\right)^n \times \left(\frac{1}{4}(\lambda^2+16)\sigma^2\right)^n > 1$.

Remark 1.3.34. *In particular, from this, we notice that if at $t = 0$ we are a-equidispersed, an instant after we are abc-overdispersed, in fact:*

$$\mathcal{I}_M^{abc} = \left(\frac{1}{4}\lambda^2\sigma^2\right)^n \times \left(\frac{1}{4}(\lambda^2+16)\sigma^2\right)^n = (1+4\sigma^2)^n > 1.$$

This is in agreement with the dispersive properties of the family $\mathcal{F}$ and legitimates, in some sense, our choice of Indexes of Dispersion*. Moreover, if the*

$$\mathcal{I}_M^{abc} = (1+4\sigma^2)^n > 1,$$

is actually different from $\mathcal{I}_M^{abc} = 5^n$, namely $\sigma^2 = 1$, we can argue that the Gaussian distribution that we are analysing does not come from the Maximal Strichartz Family of Gaussian Distributions*.*

Remark 1.3.35. *This index is different from the* Fisher Index *which is basically the variance to mean ratio*

$$\mathcal{I}_F := \frac{Var(X)}{E[X]}$$

and it is the natural one for count data. The index $\mathcal{I}_F$ is then more appropriate for families of distributions related to the Poisson distribution and that are dimensionless. In fact, in our case and in contrast with the Poisson case, we scale the Variance-Covariance Matrix *as the square of the* Expected Value*: $\Sigma \simeq \mu^2$.*

Remark 1.3.36. *The characterization of the Gaussian Distribution given by Theorem 1.3.6 and Theorem 1.2.9 can be used also to give a measure of dispersion with respect to the* Maximal Family of Gaussian Distributions*, considering the Strichartz Norm:*

$$\mathcal{I}_S := \frac{||e^{it\Delta}u_0(x)||_{L_t^q L_x^r}}{S(n,r)}.$$

By Theorem 1.3.6 one has that $0 \leq \mathcal{I}_S \leq 1$. *When the index is close to one, the distribution is close, in some sense, to the family* $\mathcal{F}$, *while, when the index is close to zero, the distribution is very far from* $\mathcal{F}$. *This index clearly does not distinguish between distributions in the family* $\mathcal{F}$. *It would be very interesting to see if the closeness to one of the index of dispersion* $\mathcal{I}_S$ *computed on a general distribution implies a proximity to the* Maximal Family of Gaussian Distributions *also from the distribution point of view and not just from the point of view of the dispersion.*

1.3.5 Partial Stochastic Ordering on $\mathcal{F}$

Using the concept of *Index of Dispersion*, we can give a *Partial Stochastic Order* to the family $\mathcal{F}$. For a more complete treatment on *Stochastic Orders*, we refer to [76]. We start the analysis of this section with the definition of *Mean-Preserving Spread*.

Definition 1.3.37. *A* mean-preserving spread *(MPS) is a map from* $\mathcal{P}(\mathbb{R}^n)$ *to itself*

$$p(x;\theta_1) \rightarrow p(x;\theta_2)$$

where $p(x;\theta_1)$ *and* $p(x;\theta_2)$ *are respectively the pdf of the random variables* X_1 *and* X_2 *with the property of leaving the* Expected Value *unchanged:*

$$\mu_{X_1}(\theta_1) = \mu_{X_2}(\theta_2),$$

for any θ_1 *and* θ_2 *in the space of parameters.*

The concept of a *Mean-Preserving Spread* provides a partial ordering of probability distributions according to their level of dispersion. We then give the following definition.

Definition 1.3.38. *Consider two random variables* X_1 *and* X_2 *such that* $\mu_{X_1}(\theta_1) = \mu_{X_2}(\theta)$, *for any* θ_1 *and* θ_2. *We say that the two random variables are* Ordered accordingly to their Dispersion Index $\mathcal{I}$ *if and only if the following condition is satisfied*

$$X_1 \prec X_2 \Leftrightarrow \mathcal{I}(X_2) \leq \mathcal{I}(X_1).$$

Now, we give some examples of ordering accordingly to the indexes of dispersion that we discussed above.

- In the case $t = 0$, we choose $||\Sigma||_a = det(\Sigma)$ and so we get

$$\mathcal{I}_M^a = ||\Sigma||_a = det(\Sigma) = \left(\frac{1}{4}\lambda^2\right)^n det(R^T R).$$

Now, in the spherical case $R^T R = \sigma^2 Id$, one gets

$$\mathcal{I}_M^a = \left(\frac{1}{4}\lambda^2\sigma^2\right)^n.$$

Using this index, we have the following partial-ordering

$$X_1 \prec X_2 \Leftrightarrow \lambda_2^2\sigma_2^2 \leq \lambda_1^2\sigma_1^2.$$

This order does not depend on the dimension n. By choosing instead $||\Sigma||_a = tr(\Sigma)$, we obtain:

$$\mathcal{I}_M^a = ||\Sigma||_a = tr(\Sigma) = \left(\frac{1}{4}\lambda^2\right) tr(R^T R).$$

Now, again in the spherical case $R^T R = \sigma^2 Id$, one gets

$$X_1 \prec X_2 \Leftrightarrow \lambda_2^2\sigma_2^2 \leq \lambda_1^2\sigma_1^2,$$

which is the same ordering as before. This order does not depend on the dimension n again and this seems to suggest that even if the value of the *Dispersion Index* might depend on the choice of the norms, the *Partial Order* is less sensible to it.

Remark 1.3.39. *In the strictly Strichartz case $\sigma^2 = 1$, we have that the* Stochastic Order *is given just by the scaling factor λ.*

- In the case when t is different from zero, we have:

$$\Sigma(\mu) = \frac{1}{4}\left(\lambda^2 + 16\Big(\frac{(\mu - x_0)\cdot v_0}{|v_0|^2}\Big)^2\right) R^T R.$$

If now we choose $x_0 = 0$ and $\mu = v_0 = (1, 0, \ldots, 0)^T$, we get

$$\Sigma(\mu) = \frac{1}{4}\Big(\lambda^2 + 16\Big) R^T R,$$

so, for $||\Sigma||_a = ||\Sigma||_b = det(\Sigma)$ and $||\mu||_c^2 = \Sigma_{1=1}^n |\mu_i|^2$, we get, in the spherical case $R^T R = \sigma^2 Id$, the following *Partial Order*:

$$X_1 \prec X_2 \Leftrightarrow \left(\frac{1}{4}\lambda_2^2\sigma_2^2\right)^n \times \left(\frac{1}{4}(\lambda_2^2 + 16)\sigma_2^2\right)^n \leq \left(\frac{1}{4}\lambda_1^2\sigma_1^2\right)^n \times \left(\frac{1}{4}(\lambda_1^2 + 16)\sigma_1^2\right)^n.$$

Remark 1.3.40. *Again, in the strictly Strichartz case $\sigma^2 = 1$, we have that the* Stochastic Order *is given just by the scaling factor λ.*

Remark 1.3.41. *In the case of the the* a-Static Dispersion Index of the Maximal Family of Gaussians $\mathcal{I}^a_M$, *the role of* σ^2 *and* λ^2 *seem interchangeable. This suggests a dimensional reduction in the parameter space, but, when* $t \neq 0$, σ^2 *and the parameter* λ *decouple and start to play a slightly different role. This suggest again a way to distinguish between* Gaussian Distributions *which come from the family* $\mathcal{F}$ *and Gaussians which do not, and so to distinguish between Gaussians which are solutions of the* Linear Schrödinger Equation, *and Gaussians which are not.*

Remark 1.3.42. *Using the definition of* Entropy, *we deduce that, for* Gaussian Distributions, $h(X) = \frac{1}{2}\log(2\pi e)^n det(\Sigma)$. *We see that, for our family* $\mathcal{F}$, *the* Entropy increases, *every time that we increase* λ, σ^2, t, *but not when we increase* x_0 *and* v_0. *In particular, the fact that the* Entropy *increases with* t *is in accordance with the* Second Principle of Thermodynamics.

Remark 1.3.43. *It seems that, the construction of similar indexes, can be performed in more general situations. In particular, we think that an index similar to* $\mathcal{I}^{abc}_M$ *can be computed in every situation in which a family of distributions has the* Variance-Covariance Matrix *and the* Expected value *which depend on common parameters.*

1.4 Strichartz Estimates as an Optimization Problem

In this section, we rephrase the problem of finding the maximizer of the Strichartz Estimates as an optimization problem. We start with the Projection Method and then we discuss the Newton Method. We refer to [19] for more details on these methods.

1.4.1 The Optimization Algorithm for the Projection Method

The scheme that we will use here is is an *iterative ascent algorithm* combined with a *projection approach*. A detailed description of the method can be found in [7] and the reference therein.

We concentrate in the case of spatial dimension $n = 1$ and critical exponent $p = 6$. The cost functional is given by:

$$J(u_0) := \sup_{u_0 \neq 0, u_0 \in L^2_x(\mathbb{R}^n)} \|e^{i\Delta} u_0\|^6_{L^6_{t,x}(\mathbb{R}\times\mathbb{R}^n)}.$$

Now, we need to compute the gradient of our cost functional J. We do not have any Lagrange multiplier, since the constraint will be later imposed by projection. The Gatauex derivative of $J(u_0)$ can be computed as follows. We have

$$J(u_0+\delta h_0)=\int_{\mathbb{R}^n\times\mathbb{R}^+}|e^{it\Delta}(u_0+\delta h_0)|^6.$$

A trick here is to think about $|u|^{2p}$ as $(u\bar{u})^p$. We get:

$$\frac{\partial}{\partial\delta}J(u_0+\delta h_0)=$$

$$3\int_{\mathbb{R}^n\times\mathbb{R}^+}|e^{it\Delta}(u_0+\delta h_0)|^4(2Re(\bar{u}h_0)+2\delta|e^{it\Delta}h_0|^2)$$

and so

$$J'(u_0)=\frac{\partial}{\partial\delta}|_{\delta=0}J(u_0+\delta h_0)=6\int_{\mathbb{R}^n\times\mathbb{R}^+}|e^{it\Delta}u_0|^4Re(\bar{u}h_0).$$

The optimization of the functional J is constrained by the flow of the PDE, namely

$$\begin{cases} iu_t(t,x)+\Delta u(t,x)=0,\ x\in\mathbb{R}^n,t\in\mathbb{R}^+\\ u(0,x)=u_0(x).\end{cases}\tag{1.37}$$

and

$$\begin{cases} ih_t(t,x)+\Delta h(t,x)=0,\ x\in\mathbb{R}^n,t\in\mathbb{R}^+\\ h(0,x)=h_0(x).\end{cases}\tag{1.38}$$

We introduce the adjoint variable u^* and integrate it against the equation for h. We get:

$$0=Re\int_{\mathbb{R}^n\times\mathbb{R}^+}(ih_t+\Delta h)\bar{u}^*=Re\int_{\mathbb{R}^n\times\mathbb{R}^+}ih_t\bar{u}^*+Re\int_{\mathbb{R}^n\times\mathbb{R}^+}\Delta h\bar{u}^*=$$

$$Re\ i\int_{\mathbb{R}^n}[h(+\infty)\bar{u}^*(+\infty)-h(0)\bar{u}^*(0)]+Re\int_{\mathbb{R}^n\times\mathbb{R}^+}(iu_t^*+\Delta u^*)\bar{h}.$$

This gives us a suggestion for the adjoint equation and so we impose:

$$\begin{cases} iu_t^*+\Delta u^*=|u|^4u,\ x\in\mathbb{R}^d,t\in\mathbb{R}\\ u^*(+\infty,x)=0.\end{cases}\tag{1.39}$$

This gives us the following condition:

$$0=-Re\ i\int_{\mathbb{R}^n}h(0)\bar{u}^*(0)+Re\int_{\mathbb{R}^n\times\mathbb{R}^+}|u|^4u\bar{h}.$$

Using this in the equation $< J'(u_0, \lambda), h_0 >= \int_{\mathbb{R}^n \times \mathbb{R}^+} |u|^4 u\bar{h}$, we obtain:

$$Re\ i \int_{\mathbb{R}^n \times \mathbb{R}^+} h_0 \bar{u}^*(0) =< J'(u_0), h_0 >$$

that can be rewritten as

$$< -iu^*(0), h_0 >_{L^2(\mathbb{R}^n)} =< J'(u_0), h >,$$

this must be true for every $h_0 \in L^2(\mathbb{R}^n)$ and so it implies that

$$-iu_0^* = J'(u_0).$$

The general strategy is to follow the algorithm described in the next few lines. For a given set of values of the mass $\|u(t)\|_{L^2(\mathbb{R}^n)} = M_0$

- Set $n = 0$ and define the tolerance ϵ.
- Set initial guess for control variable u_0.
- Solve directly the linear Schrödinger Equation.

  $$iu_t + \frac{1}{2}\Delta u = 0$$

  for initial condition $u(0, x) = u_0(x)$ has as unique solution:

  $$u(t, x) = t^{-n/2} \int_{\mathbb{R}^n} e^{i\frac{|x-y|^2}{2t}} u_0(y) dy.$$

- Solve the adjoint problem, which is

  $$\begin{cases} iu_t^* + \Delta u^* = |u|^4 u, \ x \in \mathbb{R}^d, t \in \mathbb{R} \\ \lim_{t \to +\infty} u^*(t, x) = 0. \end{cases}$$

  Note that this can be reduced to the quadratures, just by multiplying the equation by the linear propagator $e^{-it\Delta}$ and then integrate by parts, from t to $+\infty$. Therefore, we get:

  $$u^*(t) = \int_t^{+\infty} e^{i(t-s)\Delta} |e^{is\Delta} u_0|^4 e^{is\Delta} u_0 ds.$$

  This can be seen as a condition on u_0 too.

- Calculate the gradient in the proper Hilbert space H (be careful when it is not L^2)

gives

$$\nabla J(u_0) = -iu^*(0).$$

- Calculate ascent direction using Polak-Ribiére formula $p^{(n)}$. Now, we have to minimize our functional J over the manifold:

$$\mathcal{M}_{M_0} := \left\{ u(t) \in L^2(\mathbb{R}^n) \mid \|u(t)\|_{L^2(\mathbb{R}^n)} = M_0 \right\}.$$

Define

$$\psi(\tau) := \frac{M_0}{\|u^{(n)} + \tau p^{(n)}\|_{L^2(\mathbb{R}^n)}} \left(u^{(n)} + \tau p^{(n)} \right).$$

We update with

$$u^{(n+1)} = \psi(\tau_n), \quad u^{(0)} = u_0.$$

u_0 has to be chosen properly.

$$p^{(n)} = \nabla J^{(n)} - \beta_n p^{(n-1)}$$

and

$$\beta_n = \frac{\left(\nabla J^{(n)}, \nabla J^{(n)} - \nabla J^{(n-1)}\right)_{L^2(\mathbb{R}^n)}}{\|\nabla J^{(n-1)}\|^2_{L^2(\mathbb{R}^n)}}.$$

- Find the step size τ_n using arc minimization:

$$\tau_n = \arg\max \{J(\psi(\tau))\}$$

- Set $\phi^{(n+1)} = \psi(\tau_n)$.
- Evaluate

$$\Delta J = \frac{J(u^{(n+1)}) - J(u^{(n)})}{J(u^{(n)})}.$$

- Update $n \mapsto n+1$.
- While $\Delta J > \epsilon$ we keep going, otherwise we stop.

1.4.2 Optimization Algorithm with the Newton Method

An alternative way to set the optimization problem is by means of the Newton method. For more details on the Newton method, we refer to [19].

Consider the optimization problem

$$\sup_{u_0\neq 0, u_0\in L^2_x(\mathbb{R}^n)} \Psi(u_0) := \sup_{u_0\neq 0, u_0\in L^2_x(\mathbb{R}^n)} \frac{\|u_0\|^6_{L^6_{t,x}(\mathbb{R}\times\mathbb{R}^n)}}{\|u_0\|^6_{L^2_x(\mathbb{R}^n)}}$$

$$= \sup_{u_0\neq 0, u_0\in L^2_x(\mathbb{R}^n)} \|\frac{u_0}{\|u_0\|_{L^2_x(\mathbb{R}^n)}}\|^6_{L^6_{t,x}(\mathbb{R}\times\mathbb{R}^n)}$$

$$= \sup_{u_0\neq 0, \|u_0\|_{L^2_x(\mathbb{R}^n)}=M_0} \|u_0\|^6_{L^6_{t,x}(\mathbb{R}\times\mathbb{R}^n)}/M_0^6.$$

Therefore, the maximization problem of Ψ is equivalent to the following Euler-Lagrange problem:

$$J(u_0,\lambda) := \frac{\|u_0\|^6_{L^6_{t,x}(\mathbb{R}\times\mathbb{R}^n)}}{M_0^6} + \lambda(\|u_0\|^6_{L^2_x(\mathbb{R}^n)} - M_0^6).$$

The Gatauex derivative of $J(u_0,\lambda)$ is given by the following computation. We have

$$J(u_0+\delta h_0,\lambda) = \frac{1}{M_0^6}\int_{\mathbb{R}^n\times\mathbb{R}^+} |e^{it\Delta}(u_0+\delta h_0)|^6 + \lambda\left[\left(\int_{\mathbb{R}^n} |u_0+\delta h_0|^2\right)^3 - M_0^6\right].$$

Therefore, we get:

$$\frac{\partial}{\partial\delta} J(u_0+\delta h_0,\lambda) =$$

$$3M_0^{-6}\int_{\mathbb{R}^n\times\mathbb{R}^+} |e^{it\Delta}(u_0+\delta h_0)|^4 (2Re(\bar{u}h_0) + 2\delta|e^{it\Delta}h_0|^2)$$

$$+3\lambda(\int_{\mathbb{R}^n} |u_0+\delta h_0|^2)^2 \int_{\mathbb{R}^n} (2Re(\bar{u}_0 h_0) + 2\delta|h_0|^2)$$

and so

$$J'(u_0,\lambda) = \frac{\partial}{\partial\delta}|_{\delta=0} J(u_0+\delta h_0)$$

$$= 6M_0^{-6}\int_{\mathbb{R}^n\times\mathbb{R}^+} |e^{it\Delta}u_0|^4 Re(\bar{u}h_0) + 6\lambda\left(\int_{\mathbb{R}^n}|u_0|^2\right)^2 \int_{\mathbb{R}^n} Re(\bar{u}_0 h_0).$$

Since also $\frac{\partial}{\partial\lambda}J(u_0,\lambda) = 0$, then

$$\int_{\mathbb{R}^n} |u_0|^2 = M_0^2$$

and so

$$J'(u_0,\lambda) = \frac{\partial}{\partial\delta}|_{\delta=0} J(u_0+\delta h_0) = 6M_0^{-6}\int_{\mathbb{R}^{n+1}} |e^{it\Delta}u_0|^4 Re(\bar{u}h) + 6\lambda M_0^4 \int_{\mathbb{R}^n} Re(\bar{u}_0 h_0).$$

This is all under the constraint of the PDE:

$$\begin{cases} iu_t(t,x) + \Delta u(t,x) = 0, \;\; x \in \mathbb{R}^n, t \in \mathbb{R}^+ \\ u(0,x) = u_0(x). \end{cases} \tag{1.40}$$

and

$$\begin{cases} ih_t(t,x) + \Delta h(t,x) = 0, \;\; x \in \mathbb{R}^n, t \in \mathbb{R}^+ \\ h(0,x) = h_0(x). \end{cases} \tag{1.41}$$

Now, we introduce the adjoint variable u^* and integrate it against the equation for h. We get:

$$0 = Re \int_{\mathbb{R}^n \times \mathbb{R}^+} (ih_t + \Delta h)\bar{u}^* = Re \int_{\mathbb{R}^n \times \mathbb{R}^+} ih_t \bar{u}^* + Re \int_{\mathbb{R}^n \times \mathbb{R}^+} \Delta h \bar{u}^* =$$

$$Re\; i \int_{\mathbb{R}^n} [h(+\infty)\bar{u}^*(+\infty) - h(0)\bar{u}^*(0)] + Re \int_{\mathbb{R}^n \times \mathbb{R}^+} (iu_t^* + \Delta u^*)\bar{h}.$$

This gives us a suggestion for the adjoint equation and so we impose:

$$\begin{cases} iu_t^* + \Delta u^* = |u|^4 u, \;\; x \in \mathbb{R}^d, t \in \mathbb{R} \\ \lim_{t \to +\infty} u^*(t,x) = 0. \end{cases} \tag{1.42}$$

This gives us the following condition:

$$0 = -Re\; i \int_{\mathbb{R}^n \times \mathbb{R}^+} h(0)\bar{u}^*(0) + Re \int_{\mathbb{R}^n \times \mathbb{R}^+} |u|^4 u \bar{h}.$$

Using this in the equation $J'(u_0, \lambda) = 0$, we obtain:

$$M_0^{-6} Re\; i \int_{\mathbb{R}^n \times \mathbb{R}^+} h_0 \bar{u}^*(0) + \lambda \int_{\mathbb{R}^n} Re(\bar{u}_0 h_0) = 0$$

than can be rewritten as

$$< -iM_0^{-6} u_0^* + \lambda u_0, h_0 >_{L^2(\mathbb{R}^n)} = 0,$$

this must be true for every $h_0 \in L^2(\mathbb{R}^n)$ and so it implies that

$$-iM_0^{-6} u_0^* + \lambda u_0 = 0.$$

Note that $\lambda = \lambda(M_0)$.

Remark 1.4.1. *The analysis about the optimization problems outlined in this subsection* *deserves further investigation and needs to be completed with the actual numerical computation of the optimizer. At the moment we are writing the thesis, this is a work in progress.*

1.5 Concluding Remarks

In this chapter, we have discussed a characterization of the MVN distribution, by means of the maximization of the Strichartz Estimates. Differently from the case of the characterization through the entropy functional maximization, the one through Strichartz Estimates does not require the constraint of fixed variance. We computed the precise optimal constant for the whole range of Strichartz admissible exponents, discussed the connection of this problem to Restriction Theorems in Fourier analysis and gave some statistical properties of the family of Gaussian Distributions which maximize the Strichartz estimates. In particular, we computed the Fisher Information matrix of the family, we gave an Index of Dispersion and we proposed a Stochastic Ordering. We concluded this chapter presenting an optimization algorithm to compute numerically the maximizers. This last part deserves further development and it is object of current research. Furthermore, Strichartz estimates are available for several dispersive PDEs and there might be characterizations of other probability distributions based on some maximization procedure related to other differential equations. This also deserves further consideration.

Chapter 2

Characterization of Distributions through ideas from Optimal Transportation

In this chapter, we present some characterizations of probability distribution using some ideas coming from Optimal Transportation. We also use some of the techniques to give some estimation procedures for parameters of distributions with lack of regularity.

For the details of the theory of optimal transportation and an extended treatment of the subject, we refer the interested reader to [120].

2.1 On $\frac{1}{\alpha}$-*Characteristic Functions* and Applications to Asymptotic Statistical Inference

In this section, we give emphasis to a method to do statistical inference and to study properties of random variables, whose probability density functions (pdfs) do not possess good regularity, decay and integrability properties.

The main tool will be what we will call $\frac{1}{\alpha}$-*Characteristic Function*, a generalization of the classical *Characteristic Function* that is basically a measurable transform of pdfs. In this perspective and using this terminology, we will restate and prove theorems, such as the *Law of Large Numbers* (LLN) and the *Central Limit Theorem* (CLT) that now, after this measurable transform, apply to basically every distribution, upon the correct choice of a free parameter α.

We apply this theory to *Hypothesis Testing* and to the construction of *Confidence Intervals* for location parameters. We connect the classical parameters of a distribution to their related $1/\alpha$-counterparts, that we will call $\frac{1}{\alpha}$-*Momenta*.

We treat in detail the case of the *Multivariate Cauchy Distribution* for which we compute explicitly all the $\frac{1}{\alpha}$-*Expected Values* and $\frac{1}{\alpha}$-*Variances* in dimension $n = 1$ and for which we construct an approximate confidence interval for the location parameter μ, by means of asymptotic theorems in the $\frac{1}{\alpha}$-context.

Among the other things and to illustrate the usefulness of this point of view, we prove some new characterizations of the Poisson Distribution, the Uniform Discrete and the Uniform Continuous Distribution.

2.1.1 Introduction and Motivation

One of the most important tool in probability and statistics is the *Characteristic Function*, which is defined as follows.

Definition 2.1.1. *Consider the* Probability Space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \nu)$*. Here* $\mathbb{R}^n$ *is the* Sample Space*,* $\mathcal{B}(\mathbb{R}^n)$ *is the* Borel σ-algebra *and* ν *is a* Probability Measure *on* $\mathcal{B}(\mathbb{R}^n)$*. Then, the* Characteristic Function *of* ν *on* $\mathcal{B}(\mathbb{R}^n)$ *is the function* $\phi_\nu(t) : \mathbb{R}^n \to \mathbb{C}$ *defined as follows:*

$$\phi_\nu(t) := \int_{\mathbb{R}^n} e^{it\cdot x} \nu(dx).$$

Remark 2.1.2. *In the following, we will mainly talk about* Characteristic Functions *of a probability measure* ν *related to discrete (e.g.* $\nu(dx) = \sum_k c_k \delta_{x_k}$*) or absolutely continuous (e.g.* $\nu(dx) = f(x)dx$*) distributions.*

For several reasons, this tool is fundamental in distribution theory and statistical inference. For example, one can compute *Momenta*, whenever they exist, just by a simple derivation procedure.

Definition 2.1.3. *Suppose* $\phi_X(t)$ *is* k *times continuously differentiable on* $\mathbb{R}^n$*. Then, we can define the* Moment of order k *by the formula*

$$E[X^k] := (-i)^k \phi_X^{(k)}(0).$$

Here $X^k = X_1^{k_1} X_2^{k_2} \dots X_n^{k_n}$ *uses the multi-index notation with* $|k| = k_1 + k_2 + \cdots + k_n$ *and* $0 \leq k_i \leq |k|$ *for every* $i = 1, \dots, n$.

Thanks to the fact that the *Characteristic Function* characterizes the distribution and thanks to the *Lévy's Continuity Theorem* (see Section 2.1.5), several important theorems can be proved by means of the *Characteristic Function*, like the *Law of Large Numbers* and the *Central Limit Theorem.* Here are the precise statements of these theorems.

Theorem 2.1.4 (Weak Law of Large Numbers)**.** *Let* $X_1, X_2, \dots$ *be iid random variables with* $E[X_i] = \mu$ *and* $Var[X_i] = \sigma^2 < +\infty$. *Define* $\overline{X}_n := \frac{1}{n}\Sigma_{i=1}^n X_i$. *Then, for every* $\varepsilon > 0$,

$$\lim_{n \to +\infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1,$$

namely $\overline{X}_n$ Converges in Probability *to* μ.

Theorem 2.1.5 (Central Limit Theorem)**.** *Let* $X_1, X_2, \dots$ *be a sequence of iid random variables with* $E[X_i] = \mu$ *and* $Var[X_i] = \sigma^2 < +\infty$. *Define* $\overline{X}_n := \frac{1}{n}\Sigma_{i=1}^n X_i$. *Suppose* $G_n(x) := P_{Y_n}(Y_n \leq x)$ *with* $Y_n := \sqrt{n}\left(\overline{X}_n - \mu\right)/\sigma$. *Then, for every* $-\infty < x < +\infty$,

$$\lim_{n \to +\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx,$$

namely $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ Converges in Distribution *to a* Standard Normal *as* $n \to +\infty$:

$$\sqrt{n}(\overline{X}_n - \mu)/\sigma \to^d \mathcal{N}(0, 1), \quad n \to +\infty.$$

These theorems are building blocks of Statistical Inference, especially for what concerns the *Asymptotic Estimation of Parameters* and for what concerns the determination of *Asymptotic Confidence Intervals* and of *Asymptotic Rejection Regions* for *Hypotheses Testing.*

A limitation of these theorems is that they do not apply to every distribution, because not all distributions possess good regularity, decay and integrability properties in order to admit at least one momentum with finite value.

The scope of this section is to extend these building blocks-theorems to distributions which do not admit the traditional momenta and to use them to do statistical inference. Therefore, we will describe and prove some theorems about something that we will call $\frac{1}{\alpha}$-*Characteristic Function* (see Section 2.1.2 below for the precise definition). This tool allows us to obtain some weak versions of the *Law of Large Numbers* (LLN) and of the

Central Limit Theorem (CLT) that can be applied to every distribution, both discrete and continuous, upon the correct choice of a free parameter α (see Section 2.1.5 below for the precise statements).

Remark 2.1.6. *We underline here that the* $\frac{1}{\alpha}$-Characteristic Function *can be re-seen in the context of measurable transforms between pdfs and that the versions of the LLN and CLT, that we will state and prove from this point of view, are essentially well known. We will give more details about this in Section 2.1.2.*

Remark 2.1.7. *We specify here that the* $\frac{1}{\alpha}$-Characteristic Function *reduces to the usual* Characteristic Function *in the case* $\alpha = 1$ *and so one can recover classical results in the case* $\alpha = 1$.

Remark 2.1.8. *The main use of the* $\frac{1}{\alpha}$-Characteristic Function *is intended to be for distributions which do not fall in the hypotheses of classical theorems (and asymptotic theorems).*

We apply this theory to *Hypothesis Testing* and to the construction of *Confidence Intervals* for location parameters. We relate the classical parameters of a distribution to their $1/\alpha$-counterparts that we will call $\frac{1}{\alpha}$-*Momenta*.

We treat in detail the case of the *Multivariate Cauchy Distribution* for which we compute explicitly all the $\frac{1}{\alpha}$-*Expected Values* and $\frac{1}{\alpha}$-*Variances* in dimension $n = 1$ and for which we construct an approximate confidence interval for the location parameter μ, by means of our asymptotic theorems in the $\frac{1}{\alpha}$-context.

Among the other things and to illustrate the properties of the $\frac{1}{\alpha}$-*Characteristic Function*, we prove some new characterizations of the Poisson Distribution, the Uniform Discrete and the Uniform Continuous Distribution.

The remaining part of the section is organized as follows. In Subsection 2.1.2, we define the $\frac{1}{\alpha}$-*Probability of an event* E, the $\frac{1}{\alpha}$-*Characteristic Function* and the $\frac{1}{\alpha}$-*Momentum of order* k. Moreover, we explain how our point of view has a simple interpretation as measurable transform of pdfs. In Subsection 2.1.3, we compute the $\frac{1}{\alpha}$-*Characteristic Functions* of some common distributions, like the *Geometric Distribution*, the *Multivariate Normal Distribution* and the *Exponential Distribution*. In Subsection 2.1.4, we prove, giving emphasis to our definitions, some Inequalities and Identities, like the *Chebyshev Inequality*, a characterization of the *Poisson Distribution* and a characterization of both the

Discrete and the *Continuous Uniform Distributions*, using the concept of $\frac{1}{\alpha}-$*Probability*. In Subsection 2.1.5, we define the notions of *Convergence in* $\frac{1}{\alpha}$*-Probability* and *Convergence in* $\frac{1}{\alpha}$*-Distribution* and then use *Lévy's Continuity Theorem* to prove a generalized version of the *Law of Large Numbers* and the *Central Limit Theorem*. In Subsection 2.1.6, we use the results of Subsection 2.1.5 and *Slutsky's Theorem* to construct *Approximate Confidence Intervals* and *Approximate Rejection Regions* for *Hypothesis Testing*. In Subsection 2.1.7, we discuss the case of the *Multivariate Cauchy Distribution* and do some inference on its location parameter, using the theory of $\frac{1}{\alpha}$*-Characteristic Functions*.

Remark 2.1.9. *The characterizations of the distributions are new, as their applications in later chapters, while the asymptotic theorems that we use, state and prove, are just a little variation of essentially well known results.*

Remark 2.1.10. *We believe that several other characterizations and theorems can be proved by means of the* $\frac{1}{\alpha}$-Characteristic Function*, but we leave these for future studies.*

2.1.2 Definition of $\frac{1}{\alpha}$-*Characteristic Functions*

In this subsection, we fix some notation and we introduce the tool of $\frac{1}{\alpha}$*-Characteristic Functions*. The $\frac{1}{\alpha}$*-Characteristic Function* is basically the Fourier Transform, but, differently from the *Characteristic Function*, the average is not taken with respect to the pdf, but with respect to the $\frac{1}{\alpha}$-power of the pdf.

In the next paragraphs, we will always consider the *Probability Space* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \nu)$, where $\mathbb{R}^n$ is the *Sample Space*, $\mathcal{B}(\mathbb{R}^n)$ is the *Borel* σ*-algebra* and ν is a *Probability Measure* on $\mathcal{B}(\mathbb{R}^n)$. We will often use the notation $\nu^{1/\alpha}(dx)$ with which we will mean $\nu^{1/\alpha}(dx) = f(x)^{1/\alpha}dx$ with $x \in \mathbb{R}^n$ in the absolutely continuous case and $\nu^{1/\alpha}(dx) = \sum_k c_k^{1/\alpha}\delta_{x_k}$ with $x_k \in \mathbb{R}^n$, $k \in \mathbb{N}$ in the discrete case. When there will not be any problem of misunderstanding, we will not specify with respect to which measure we are computing the probability of an event.

We now define what we mean with $\frac{1}{\alpha}$*-Probability*.

Definition 2.1.11 ($\frac{1}{\alpha}$*-Probability of an Event* E)**.** *Consider a random variable* X *with probability measure* $\nu(dx)$*. We call* $\frac{1}{\alpha}$-Probability of an Event E*, the following quantity*

$$P^{1/\alpha}(X \in E) := \frac{\int_E \nu^{1/\alpha}(dx)}{\int_{\mathbb{R}^n} \nu^{1/\alpha}(dx)}.$$

With this definition, we can specify what we mean with $\frac{1}{\alpha}$-*Momentum of order k*.

Definition 2.1.12 ($\frac{1}{\alpha}$-*Momentum of order k*)**.** *Consider a random variable X with probability measure $\nu(dx)$. We call $\frac{1}{\alpha}$-*Momentum of order k, *the following quantity*

$$E^{1/\alpha}(X^k) := \frac{\int_E x^k \nu^{1/\alpha}(dx)}{\int_{\mathbb{R}^n} \nu^{1/\alpha}(dx)}.$$

Here $X^k := X_1^{k_1} X_2^{k_2} \dots X_n^{k_n}$ and $x^k := x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}$ use the multi-index notation with $|k| = k_1 + k_2 + \dots + k_n$ and $0 \leq k_i \leq |k|$ for every $i = 1, \dots, n$.

Remark 2.1.13. *These quantities are new characteristic values of each particular distribution and characterize it, as much as classical momenta do.*

We are ready to define the notion of $\frac{1}{\alpha}$-*Characteristic Function*.

Definition 2.1.14 ($\frac{1}{\alpha}$-*Characteristic Function*)**.** *Consider a random variable X with probability measure $\nu(dx)$. Then, the* $\frac{1}{\alpha}$-Characteristic Function *of ν on $\mathcal{B}(\mathbb{R}^n)$ is the function $\phi_X^{\frac{1}{\alpha}}(\xi) : \mathbb{R}^n \to \mathbb{C}$ defined as follows:*

$$\phi_X^{\frac{1}{\alpha}}(\xi) := \frac{\int_{\mathbb{R}^n} e^{-ix\cdot\xi} \nu^{1/\alpha}(dx)}{\int_{\mathbb{R}^n} \nu^{1/\alpha}(dx)}.$$

Proposition 2.1.15. *Consider $\phi_X^{\frac{1}{\alpha}}(\xi)$ the* $\frac{1}{\alpha}$-Characteristic Function *of a random variable X Then, if $\phi_X^{\frac{1}{\alpha}}(\xi)$ is k times continuously differentiable on $\mathbb{R}$, we have that the* $\frac{1}{\alpha}$-Momentum of order k *can be computed by the formula*

$$E^{\frac{1}{\alpha}}[X^k] := (-i)^k (\phi_X^{\frac{1}{\alpha}})^{(k)}(0).$$

Proof. It is a direct and simple computation. □

Remark 2.1.16. *The concepts that we defined in this section can be interpreted also as follows. Consider X a random variable. Then, we can define X^h, a new random variable, in the following way:*

$$P(X^h \in E) := E[\mathbf{1}(X \in E)h(X)/c],$$

provided that h is a measurable transform and $c \in (0, \infty)$ is given by $c := E[h(X)]$. In our case, $h(t)$ is chosen to be $h(t) = f^{1/\alpha - 1}(t)$. This point of view is legitimate and simplifies the treatment of the asymptotic theorems in later sections. We decided to keep the point of view of the $1/\alpha$-Characteristic Function, *because, in our opinion, it better explains what the transform does to the pdf, namely a redistribution of its mass. Moreover, the $1/\alpha$-point* *of view connects more directly to the analysis of the* Consensus Monte Carlo Algorithm *for* Big Data *introduced in [105], as we will explain in detail in Section 2.3.*

2.1.3 The $\frac{1}{\alpha}$-*Characteristic Functions* of some common distributions

In this subsection, we compute the $\frac{1}{\alpha}$-*Characteristic Function* of some common distributions. In particular we compute the $\frac{1}{\alpha}$-*Characteristic Functions* of the *Geometric Distribution*, the *Multivariate Normal Distribution* and the *Exponential Distribution*.

Remark 2.1.17. *We note first that for some common distributions, like the Binomial, the Poisson, the Negative Binomial and all the distributions whose pdf contains some factorials (so mainly distributions which, in some way, involve some counting), the* $\frac{1}{\alpha}$-Characteristic Function *is not available in explicit form. Nevertheless, for these distributions it is already possible to construct all the* Momenta *that you need and so one does not really need the explicit form of the* $\frac{1}{\alpha}$-Characteristic Function.

The main attention is then posed on distributions for which we can already talk about the usual *Momenta* and we will leave for a later section the discussion on the *Cauchy Distribution*, that does not admit any *Momenta*.

- *The Geometric Distribution* $f(k;p) = (1-p)^k p$ for $k = 0, 1, 2, \dots$ and $0 < p < 1$. The $\frac{1}{\alpha}$-*Characteristic Function* can be computed in the following way.

  $$\phi_X^{\frac{1}{\alpha}} = \frac{NUM}{DEN}$$

  and so

  $$NUM = \Sigma_{k=0}^{+\infty} e^{-ik\xi}(1-p)^{k/\alpha} p^{1/\alpha}.$$

  Now take $q = (1-p)^{1/\alpha}$ and so $p^{1/\alpha} = (1-q^\alpha)^{1/\alpha}$, therefore

  $$NUM = (1-q^\alpha)^{1/\alpha} \Sigma_{k=0}^{+\infty} (e^{-ik\xi} q)^k.$$

  Now, since $0 < p < 1$, also $0 < q < 1$ and therefore $(e^{-ik\xi}q)^k$ is the general term of an absolutely convergent series. Then

  $$NUM = (1-q^\alpha)^{1/\alpha} \frac{1}{1 - e^{-i\xi} q}.$$

  To obtain the normalizing constant, one needs to substitute, $\xi = 0$ into the previous formula and so get

$$\phi^{1/\alpha}(\xi) = NUM/DEN = \frac{1-(1-p)^{1/\alpha}}{1-(1-p)^{\frac{1}{\alpha}}e^{-i\xi}}$$

- *The Multivariate Normal Distribution* The pdf of the *Multivariate Normal Distribution* can be written in the following way:

$$f_{\mathbf{x}}(x_1,\ldots,x_n) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{\Sigma}|}}\exp\left(-\frac{1}{2}(x-\mu)^T\mathbf{\Sigma}^{-1}(x-\mu))\right),$$

where $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$ and $|\mathbf{\Sigma}|$ is the determinant of $\mathbf{\Sigma}$, which is a positive definite symmetric $(n\times n)$-matrix. Again, the $\frac{1}{\alpha}$-*Characteristic Function* can be computed in the following way:

$$\phi_X^{\frac{1}{\alpha}} = \frac{NUM}{DEN}$$

and so

$$\begin{aligned}
DEN = \quad & \int_{\mathbb{R}^n}\left(\frac{1}{\sqrt{(2\pi)^n|\mathbf{\Sigma}|}}\right)^{1/\alpha}\exp\left(-\frac{1}{2\alpha}(x-\mu)^T\mathbf{\Sigma}^{-1}(x-\mu)\right)dx \\
= \quad & \frac{\alpha^{n/2}|\Sigma|^{1/2-1/2\alpha}}{(2\pi)^{n/2\alpha}}\int_{\mathbb{R}^n}e^{-\frac{1}{2}|y|^2}dy = \alpha^{n/2}|\Sigma|^{1/2-1/2\alpha}(2\pi)^{n/2-n/2\alpha}.
\end{aligned}$$

by the change of variables $y = \alpha^{-1/2}\Sigma^{-1/2}(x-\mu)$.

$$\begin{aligned}
NUM = \quad & \int_{\mathbb{R}^n}e^{-i\xi\cdot x}\left(\frac{1}{\sqrt{(2\pi)^n|\mathbf{\Sigma}|}}\right)^{1/\alpha}\exp\left(-\frac{1}{2\alpha}(x-\mu)^T\mathbf{\Sigma}^{-1}(x-\mu)\right)dx \\
& = \alpha^{n/2}|\Sigma|^{1/2-1/2\alpha}(2\pi)^{-n/2\alpha}e^{-i\mu\cdot\xi}\int_{\mathbb{R}^n}e^{-\frac{1}{2}|y|^2-i\alpha^{1/2}\Sigma^{1/2}y\cdot\xi}dy \\
& = \alpha^{n/2}|\Sigma|^{1/2-1/2\alpha}(2\pi)^{-n/2\alpha}e^{-i\mu\cdot\xi-\frac{1}{2}\alpha\xi^T\Sigma\xi} \\
\times \quad & \int_{\mathbb{R}^n}e^{-\frac{1}{2}\left(|y|^2-2i\alpha^{1/2}\Sigma^{1/2}y\cdot\xi-\alpha\xi^T\Sigma\xi\right)}dy \\
& = \alpha^{n/2}|\Sigma|^{1/2-1/2\alpha}(2\pi)^{-n/2\alpha}e^{-i\mu\cdot\xi-\frac{1}{2}\alpha\xi^T\Sigma\xi}\int_{\mathbb{R}^n}e^{-\frac{1}{2}|z|^2}dz \\
& = \alpha^{n/2}|\Sigma|^{1/2-1/2\alpha}(2\pi)^{n/2-n/2\alpha}e^{-i\mu\cdot\xi-\frac{1}{2}\alpha\xi^T\Sigma\xi}
\end{aligned}$$

by the change of variables $y = \alpha^{-1/2}\Sigma^{-1/2}(x-\mu)$ and $y = z - i\alpha^{1/2}\Sigma^{1/2}\xi$. Therefore:

$$\phi^{1/\alpha}(\xi) = NUM/DEN = e^{-i\mu\cdot\xi - \frac{1}{2}\alpha\xi^T\Sigma\xi}.$$

- *The Exponential Distribution* $f(x;\lambda) = \lambda e^{-\lambda x}$ for $x > 0$

  Similarly as before, the $\frac{1}{\alpha}$-*Characteristic Function* can be computed in the following way:

$$\phi_X^{\frac{1}{\alpha}} = \frac{NUM}{DEN}$$

  and so

$$DEN = \int_0^{+\infty} \lambda^{1/\alpha} e^{-x\lambda/\alpha} dx = \lambda^{1/\alpha - 1}\alpha \int_0^{+\infty} e^{-y} dy = \lambda^{1/\alpha-1}\alpha, \tag{2.1}$$

  by the change of variable $y = x\lambda/\alpha$.

$$NUM = \int_0^{+\infty} \lambda^{1/\alpha} e^{-ix\xi - x\lambda/\alpha} dx = \frac{\lambda^{1/\alpha}}{\lambda/\alpha + i\xi} \int_0^{+\infty} e^{-y} dy = \frac{\lambda^{1/\alpha}}{\lambda/\alpha + i\xi} \tag{2.2}$$

  by the change of variable $y = x\lambda/\alpha + ix\xi$. So the $\frac{1}{\alpha}$-*Characteristic Function* is

$$\phi^{1/\alpha}(\xi) = NUM/DEN = \frac{\alpha^{-1}\lambda^{1/\alpha+1-1/\alpha}}{\lambda/\alpha + i\xi} = \frac{\lambda}{\lambda + i\alpha\xi}. \tag{2.3}$$

Remark 2.1.18. *We notice that, in all the above cases, we recover the usual* Characteristic Function *if we choose $\alpha = 1$ and this is a general fact and very easy to prove. Also, we notice that the Gaussian is a fixed point of the* $\frac{1}{\alpha}$-Characteristic Function *(up to constants) as it is for the usual* Characteristic Function*.*

Remark 2.1.19. *The $P^{\frac{1}{\alpha}}(X = x)$ of discrete and continuous distributions does not necessarily coincide with $P(X = x)$. We refer to Section 2.1.4, for a characterization of distributions for which this fact does occur.*

2.1.4 Identities and Inequalities

In this subsection, we prove some identities and inequalities, which will be useful for the next section. In particular, we prove a *Chebyshev Inequality*, a characterization of the *Poisson Distribution* and a characterization of both the *Discrete* and the *Continuous*

Uniform Distributions, using the concept of $\frac{1}{\alpha}-$*Probability*. Some of these results are of independent interest and serve also for the sake of illustration of the theory.

Remark 2.1.20. *A lot of other identities and inequalities can be proved in a similar way to the case $\alpha = 1$. We refer to [25] for a more complete set of references and type of inequalities that one can try to reprove in this context. We plan to discuss them in a future work.*

We start with the *Chebyshev Inequality*.

Theorem 2.1.21 (Chebyshev Inequality)**.** *Let X be a random variable and $g(x) \geq 0$. Then, for any $r > 0$, the following inequality holds:*

$$P^{\frac{1}{\alpha}}(g(X) \geq r) \leq \frac{E^{\frac{1}{\alpha}}[g(X)]}{r}. \tag{2.4}$$

Proof. To fix the ideas, we prove the theorem in the continuous case. The discrete one follows similarly. We just need to perform a series of inequalities. Let $c_\alpha^{-1} := \int_{-\infty}^{+\infty} f(x)^{\frac{1}{\alpha}} dx$ be the normalizing constant.

$$E^{\frac{1}{\alpha}}[g(X)] = c_\alpha \int_{-\infty}^{+\infty} g(x) f(x)^{\frac{1}{\alpha}} dx \geq c_\alpha \int_{\{x: g(x) \geq r\}} g(x) f(x)^{\frac{1}{\alpha}} dx \tag{2.5}$$

$$\geq c_\alpha \int_{\{x: g(x) \geq r\}} r f(x)^{\frac{1}{\alpha}} dx = r P^{\frac{1}{\alpha}} \left(X \in \{x : g(x) \geq r\}\right) = r P^{\frac{1}{\alpha}} \left(g(X) \geq r\right). \tag{2.6}$$

This concludes the proof of the theorem. □

Remark 2.1.22. *The proof does not require the probability context (in fact c_α does not play any role), so the inequality works in the more general setting of measure theory and can be in fact seen as the classical* Chebyshev Inequality *applied to the random variable X^h (See Remark 2.1.16). It can be proved that* Chebyshev Inequality *is a conservative inequality, namely it is rarely attained as in the case $\alpha = 1$. For the case $\alpha = 1$ we refer to [55].*

Example 2.1.23. *We consider the case $r = t^2$ with $t > 0$ and $g(x) = (x - \mu_\alpha)^2/\sigma_\alpha^2$ where $\mu_\alpha := E^{\frac{1}{\alpha}}[X]$ and $\sigma_\alpha^2 := Var^{\frac{1}{\alpha}}[X] = E^{\frac{1}{\alpha}}[X^2] - E^{\frac{1}{\alpha}}[X]^2$. Plugging this inside the* Chebyshev Inequality*, we obtain:*

$$P^{\frac{1}{\alpha}}\left((X - \mu_\alpha)^2/\sigma_\alpha^2 \geq t^2\right) \leq \frac{E^{\frac{1}{\alpha}}[(X - \mu_\alpha)^2/\sigma_\alpha^2]}{t^2} = \frac{1}{t^2}. \tag{2.7}$$

This can be rewritten in the following two ways:

$$P^{\frac{1}{\alpha}}(|X-\mu_\alpha| \geq \sigma_\alpha t) \leq \frac{1}{t^2} \tag{2.8}$$

or

$$P^{\frac{1}{\alpha}}(|X-\mu_\alpha| \leq \sigma_\alpha t) \geq 1-\frac{1}{t^2}, \tag{2.9}$$

which is a control on the $\frac{1}{\alpha}$-Probability *tails.*

Remark 2.1.24. *The example would have not worked if instead of using the* μ_α *and* σ_α *we would have used the usual* μ *and* σ.

We pass now to a characterization of the *Poisson Distribution*.

Proposition 2.1.25 (Poisson's Identity)**.** *A random variable* $X \sim Poiss(\lambda)$, *namely its pdf is* $f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x \in \mathbb{N}$ *if and only if*

$$P^{\frac{1}{\alpha}}(X = x+1) = \left(\frac{\lambda}{x+1}\right)^{\frac{1}{\alpha}} P^{\frac{1}{\alpha}}(X = x),$$

for some $\alpha > 0$.

Proof. First, we suppose $X \sim Poiss(\lambda)$. Then, if we call the normalizing constant

$$c_\alpha := \Sigma_{x=1}^{+\infty} f(x;\lambda)^{1/\alpha},$$

we have

$$P^{\frac{1}{\alpha}}(X = x+1) = c_\alpha \frac{e^{-\lambda/\alpha}\lambda^{(x+1)/\alpha}}{(x+1)!^{1/\alpha}} = c_\alpha \frac{e^{-\lambda/\alpha}\lambda^{1/\alpha}}{(x+1)^{1/\alpha}} \frac{\lambda^{x/\alpha}}{(x!)^{1/\alpha}} = \left(\frac{\lambda}{x+1}\right)^{\frac{1}{\alpha}} P^{\frac{1}{\alpha}}(X = x).$$

On the other side, if

$$P^{\frac{1}{\alpha}}(X = x+1) = \left(\frac{\lambda}{x+1}\right)^{\frac{1}{\alpha}} P^{\frac{1}{\alpha}}(X = x),$$

then

$$P(X = x+1) = \left(\frac{\lambda}{x+1}\right) P(X = x),$$

because $P^{\frac{1}{\alpha}}(X = x) = c_\alpha P(X = x)^{\frac{1}{\alpha}}$, for every $x \in \mathbb{N}$. Now, for every $x \in \mathbb{N}$, we define

$$p_x := P(X = x)$$

and so we have:

$$p_{x+1}/p_x = \frac{\lambda}{x+1},$$

which by recursion becomes

$$p_x/p_0 = \frac{\lambda^x}{x!}.$$

Now, we use the normalizing condition to get p_0:

$$1 = \Sigma_{x=0}^{+\infty} p_0 \frac{\lambda^x}{x!} = p_0 \Sigma_{x=0}^{+\infty} \frac{\lambda^x}{x!} = p_0 e^{\lambda}$$

and so $p_0 = e^{-\lambda}$. Therefore,

$$f(x; \lambda) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!},$$

for every $x \in \mathbb{N}$ and so $X \sim Poiss(\lambda)$. This concludes the proof of the theorem. □

Here we give a characterization of the *Discrete Uniform Distribution*.

Theorem 2.1.26 (Characterization of the Discrete Uniform Distribution)**.** *For every $N \in \mathbb{N}$, the following equivalence is true. A random variable X follows the* Discrete Uniform Distribution *$X \sim Uniform(1, N)$ ($f(j, N) = 1/N$ for every $j = 1, \dots, N$) if and only if the following condition holds*

$$P^{\frac{1}{\alpha}}(X = j) = P(X = j)$$

for every $j = 1, \dots, N$.

Proof. First of all, we define $p_j := P(X = j)$ so that $P^{\frac{1}{\alpha}}(X = j) = \frac{p_j^{1/\alpha}}{\Sigma_{i=1}^{N} p_i^{1/\alpha}}$. Therefore, the condition $P^{\frac{1}{\alpha}}(X = j) = P(X = j)$ for every$j = 1, \dots, N$ is equivalent to

$$\frac{p_j^{1/\alpha}}{\Sigma_{i=1}^{n} p_i^{1/\alpha}} = p_j$$

for every $j = 1, \dots, N$. One direction is trivial to prove, since you can just plug in this last formula $p_1 = \dots = p_N = 1/N$ and the identity is satisfied. Now, we suppose that the identity is satisfied, namely that

$$P^{\frac{1}{\alpha}}(X=j)=P(X=j)$$

for every $j=1,\dots,N$ and so that $\frac{p_j^{1/\alpha}}{\Sigma_{i=1}^N p_i^{1/\alpha}}=p_j$ for every $j=1,\dots,N$. Since $\Sigma_{i=1}^N p_i=1$, at least one of the p_i's is not zero. We take $j\neq k$ with $p_k\neq 0$ and we take the ratio side by side of the previous equality, so that we have

$$\frac{p_j^{1/\alpha}}{p_k^{1/\alpha}}=\frac{\frac{p_j^{1/\alpha}}{\Sigma_{i=1}^N p_i^{1/\alpha}}}{\frac{p_k^{1/\alpha}}{\Sigma_{i=1}^N p_i^{1/\alpha}}}=\frac{p_j}{p_k}$$

for every $j=1,\dots,N$ and $k=1,\dots,N$ but $j\neq k$. So, since $p_k\neq 0$, and for $p_j\neq 0$, we have $\frac{p_j^{1/\alpha-1}}{p_k^{1/\alpha-1}}=1$ and hence $p_j=p_k$ for every j,k such that $p_j\neq 0$. So, $p_1=\cdots=p_N=\frac{1}{N}$ where N is the number of indexes j such that $p_j\neq 0$. This proves the theorem. □

Here we give a characterization of the *Continuous Uniform Distribution.*

Theorem 2.1.27 (Characterization of the Continuous Uniform Distribution)**.** *Suppose $f(t)$ is a pdf such that $F(x)=\int_0^x f(t)dt$ is differentiable in x. A random variable X with pdf $f(x)$ follows the* Continuous Uniform Distribution *if and only if $P^{\frac{1}{\alpha}}(X\in E)=P(X\in E)$, for every $E\subset\mathbb{R}^n$ measurable.*

Proof. If $X\sim U(a,a+1)$ then its pdf is $f(x)=1$ for $a\leq x\leq a+1$ and 0 otherwise. Therefore $f(x)^{\frac{1}{\alpha}}=1$ for $a\leq x\leq a+1$ and 0 otherwise. Hence, the two pdfs coincide and so $P^{\frac{1}{\alpha}}(X\in E)=P(X\in E)$. Similarly if instead of $(a,a+1)$ we consider a more general set. Now, suppose that $P^{\frac{1}{\alpha}}(X\in E)=P(X\in E)$ is true for every $E\subset\mathbb{R}^n$ measurable, so it is true in particular for $E=[c,b]$. Therefore,

$$\frac{\int_c^b f(x)^{\frac{1}{\alpha}}dx}{\int_{\mathbb{R}} f(x)^{\frac{1}{\alpha}}dx}=\int_c^b f(x)dx.$$

If we take $c=0$ and $b=x$, this implies that

$$F(x)=\frac{\int_0^x f(t)^{\frac{1}{\alpha}}dt}{\int_{\mathbb{R}} f(x)^{\frac{1}{\alpha}}dx}=\int_0^x f(t)dt=G(x)$$

and so by taking the derivative in x (which is allowed by Lebesgue Differentiation Theorem),

$$F'(x) = \frac{f(x)^{\frac{1}{\alpha}}}{\int_{\mathbb{R}} f(x)^{\frac{1}{\alpha}} dx} = f(x) = G'(x).$$

This implies that $f(x) \neq 0$ if and only if $f(x) = c_\alpha := \left(\int_{\mathbb{R}} f(x)^{\frac{1}{\alpha}} dx\right)^{\frac{\alpha}{\alpha-1}}$, and so that $X \sim U(I(c_\alpha))$. This completes the proof of the theorem. □

Remark 2.1.28. *Both these theorems have a simple adaptation to the multivariate case. Moreover, other theorems can be extended to the $\frac{1}{\alpha}$-case. We refer to [25] pages 121-127 and 187-192 for a set of Identities and Inequalities that can be reproved in the $\frac{1}{\alpha}$-context. We leave this for a future work.*

2.1.5 *Characteristic Functions* and Weak Convergence

In this subsection, we will discuss the interplay between the concepts of $\frac{1}{\alpha}$-*Characteristic Function* and *Weak Convergence*. This will allow us to prove the $\frac{1}{\alpha}$-counterpart of some very famous and important theorems, like the *Law of Large Numbers* and the *Central Limit Theorem*. Again, this theorems are just small adaptations of well known results (See Remark 2.1.16).
We first recall *Lévy's Continuity Theorem*, which will turn out to be useful in the following.

Theorem 2.1.29 (Lévy's Continuity Theorem)**.** *Let $(\nu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathbb{R}^n$, with* Characteristic Functions $(\phi_n)_{n\in\mathbb{N}}$. *If $\nu_n \to^w \nu_\infty$, then ϕ_n converges pointwise to ϕ_∞, the* Characteristic Function *of ν_∞. Conversely if ϕ_n converges pointwise to a function ϕ_∞ which is continuous at 0, then ϕ_∞ is the* Characteristic Function *of a probability measure ν_∞ and $\nu_n \to^w \nu_\infty$.*

Proof. See for example [16]. □

2.1.5.1 The Law of Large Numbers and the Central Limit Theorem

Before stating and proving a generalization of the *Law of Large Numbers*, we have to introduce the notion of *Convergence in $1/\alpha$-Probability.*

Definition 2.1.30 (Convergence in $1/\alpha$-Probability)**.** *We say that a sequence of random variables $X_1, X_2, \dots$* Converges in $1/\alpha$-Probability *to a random variable X if, for every $\varepsilon > 0$, we have*

$$\lim_{n\to+\infty} P^{1/\alpha}(|X_n - X| < \varepsilon) = 1.$$

We are now ready to prove a generalization of the *Law of Large Numbers.*

Theorem 2.1.31 (Weak Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $E^{\frac{1}{\alpha}}[X_i] = \mu_\alpha$ and $Var^{\frac{1}{\alpha}}[X_i] = \sigma_\alpha^2 < +\infty$ for every $i \in \mathbb{N}$ and for a certain $\alpha \in \mathbb{R}$. Define $\overline{X}_n := \frac{1}{n}\Sigma_{i=1}^n X_i$. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to+\infty} P^{1/\alpha}(|X_n - \mu_\alpha| < \varepsilon) = 1,$$

namely $\overline{X}_n$ converges in $1/\alpha$-Probability *to μ_α.*

Proof. It is a straightforward application of the *Chebyshev Inequality*. For every $\varepsilon > 0$, we compute

$$P^{1/\alpha}(|\overline{X}_n - \mu_\alpha| \geq \varepsilon) = P^{1/\alpha}(|\overline{X}_n - \mu_\alpha|^2 \geq \varepsilon^2) \leq \frac{E^{1/\alpha}\left[(\overline{X}_n - \mu_\alpha)^2\right]}{\varepsilon^2} = \frac{Var^{\frac{1}{\alpha}}[\overline{X}_n]}{\varepsilon^2} = \frac{\sigma_\alpha^2}{\varepsilon^2}.$$

Therefore

$$P^{1/\alpha}(|\overline{X}_n - \mu_\alpha| < \varepsilon) = 1 - P^{1/\alpha}(|\overline{X}_n - \mu_\alpha| \geq \varepsilon) \geq 1 - \frac{\sigma_\alpha^2}{\varepsilon^2} \to 1,$$

as $n \to +\infty$ and so the theorem. □

Before stating and proving the *Central Limit Theorem*, we have to introduce the notion of *Convergence in $1/\alpha$-Distribution*.

Definition 2.1.32 (Convergence in $1/\alpha$-Distribution)**.** *We say that a sequence of random variables $X_1, X_2, \ldots$* Converges in $1/\alpha$-Distribution *to a random variable X if*

$$\lim_{n\to+\infty} F_{X_n}^{\frac{1}{\alpha}} = F_X^{\frac{1}{\alpha}}(x)$$

at all points x of continuity of $F_X^{\frac{1}{\alpha}}(x)$. Here $F_Y^{\frac{1}{\alpha}}(y) := P_Y^{1/\alpha}(Y \leq y)$.

We are now ready to prove a generalization of the *Central Limit Theorem*.

Theorem 2.1.33 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $E^{\frac{1}{\alpha}}[X_i] = \mu_\alpha$ and $Var^{\frac{1}{\alpha}}[X_i] = \sigma_\alpha^2 < +\infty$ for a certain $\alpha \in \mathbb{R}$. Suppose also that, for that $\alpha \in \mathbb{R}$, the third moment exists. Define $\overline{X}_n := \frac{1}{n}\Sigma_{i=1}^n X_i$. Suppose $G_n^{1/\alpha}(x) := P_{Y_n}^{1/\alpha}(Y_n \leq x)$ with $Y_n := \sqrt{n}\left(\overline{X}_n - \mu_\alpha\right)/\sigma_\alpha$. Then, for every $-\infty < x < +\infty$,*

$$\lim_{n\to+\infty} G_n^{1/\alpha}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx,$$

namely

$$\sqrt{n}\left(\overline{X}_n - \mu_\alpha\right)/\sigma_\alpha \to \mathcal{N}(0,1)$$

in $\frac{1}{\alpha}$-Distribution.

Proof. Consider the $\frac{1}{\alpha}$-*Characteristic Function* of $Y_n := \sqrt{n}\left(\overline{X}_n - \mu_\alpha\right)/\sigma_\alpha$, namely

$$\phi^{\frac{1}{\alpha}}_{Y_n}(\xi) = E^{\frac{1}{\alpha}}[e^{i\xi x}].$$

Moreover, consider the function $g(\xi) := log\left(\phi^{\frac{1}{\alpha}}_{Y_n}(\xi)\right)$ and expand it in ξ close to $\xi = 0$, so

$$g(\xi) = g(0) + g'(0)\xi + \frac{1}{2}g''(0)\xi^2 + o(\xi^2).$$

We now want to compute the coefficients $g(0)$, $g'(0)$ and $g''(0)$. First of all, $g(0) = log\left(\phi^{\frac{1}{\alpha}}_{Y_n}(0)\right) = log(1) = 0$. Then $g'(\xi) = \frac{\frac{d}{d\xi}\phi^{\frac{1}{\alpha}}_{Y_n}(\xi)}{\phi^{\frac{1}{\alpha}}_{Y_n}(\xi)} = iE^{\frac{1}{\alpha}}_{Y_n}[Y_n] = 0$ and then $g''(\xi) = \frac{\phi^{\frac{1}{\alpha}}_{Y_n}(\xi)\frac{d^2}{d\xi^2}\phi^{\frac{1}{\alpha}}_{Y_n}(\xi) - [\frac{d}{d\xi}\phi^{\frac{1}{\alpha}}_{Y_n}(\xi)]^2}{[\phi^{\frac{1}{\alpha}}_{Y_n}(\xi)]^2}$ and so $g''(0) = -E^{\frac{1}{\alpha}}[X^2] + E^{\frac{1}{\alpha}}[X]^2 = -Var^{\frac{1}{\alpha}}[Y_n] = -1$. This implies that

$$g(\xi) = -\frac{1}{2}g''(0)\xi^2 + o(\xi^2)$$

in a neighborhood of $\xi = 0$. Therefore,

$$\phi^{\frac{1}{\alpha}}_{Y_n}(\xi) = \phi^{\frac{1}{\alpha}}_{X_1}(\xi/\sqrt{n}) = e^{-\frac{1}{2}\xi^2 + o(\xi^3/n^{3/2})}.$$

So, pointwise for fixed ξ (and uniformly for a closed neighborhood of the origin), we have that

$$\phi^{\frac{1}{\alpha}}_{Y_n}(\xi) \to e^{-\frac{1}{2}\xi^2}, \quad \text{as } n \to +\infty.$$

By *Lévy's Continuity Theorem*, we deduce that

$$\sqrt{n}\left(\overline{X}_n - \mu_\alpha\right)/\sigma_\alpha \to \mathcal{N}(0,1)$$

in $\frac{1}{\alpha}$-*Distribution*. This completes the proof. □

Remark 2.1.34. *The theorems and notions of convergence presented above, reduce to well known results if applied to the sequence X^h_n, where X^h has been defined in Remark 2.1.16.*

Remark 2.1.35. *The request of existence of the third moment can be removed, but with the expense of complicating the proof a bit and with the benefit of improving the range of α. We decided to not pursue this direction here.*

2.1.6 Some inference: Confidence Intervals and Hypothesis Testing

In this subsection, we use the generalization of the *Law of Large Numbers* and the *Central Limit Theorem* presented above, to give, asymptotically, some estimates about μ_α and σ_α^2. We will see, in the following section, that this will indeed permit us to do some inference on classical parameters of distributions for which the classical *Law of Large Numbers* and the classical *Central Limit Theorem* do not apply. We start by discussing *Confidence Intervals* for μ_α with σ_α^2 known or unknown, to pass after, by means of the *Inversion Principle*, to *Hypothesis Testing.*

2.1.6.1 Confidence Interval for μ_α with σ_α^2 known

By the *Central Limit Theorem*, we have that

$$Y_n := \sqrt{n}(\overline{X}_n - \mu_\alpha)/\sigma_\alpha \to^d \mathcal{N}(0,1)$$

in $\frac{1}{\alpha}$-*Distribution.* Therefore, for n sufficiently big, one can consider the approximation

$$P^{\frac{1}{\alpha}}\left(\overline{X}_n - k\frac{\sigma_\alpha}{\sqrt{n}} \leq \mu_\alpha \leq \overline{X}_n + k\frac{\sigma_\alpha}{\sqrt{n}}\right) = P^{\frac{1}{\alpha}}\left(\left|\sqrt{n}(\overline{X}_n - \mu_\alpha)/\sigma_\alpha\right| \leq k\right) \simeq \Phi(k) - \Phi(-k)$$

with $\Phi(\cdot)$ the cdf of the *Standard Normal Distribution* $\mathcal{N}(0,1)$, and then impose

$$1 - \gamma = \Phi(k) - \Phi(-k)$$

to get

$$k_\gamma = \Phi^{-1}\left(1 - \frac{\gamma}{2}\right).$$

Therefore, we say that

$$\overline{X}_n - k_\gamma\frac{\sigma_\alpha}{\sqrt{n}} \leq \mu_\alpha \leq \overline{X}_n + k_\gamma\frac{\sigma_\alpha}{\sqrt{n}}$$

is an *Asymptotic Confidence Interval for* μ_α *at level* $1-\gamma$ with known $\frac{1}{\alpha}$-*Variance*, equal to σ_α^2.

Remark 2.1.36. *We underline here that, since the type of convergence depends on* α*, also the precision of the approximation*

$$P^{\frac{1}{\alpha}}\left(\left|\sqrt{n}(\overline{X}_n - \mu_\alpha)/\sigma_\alpha\right| \leq k\right) \simeq \Phi(k) - \Phi(-k)$$

varies with α*. When needed, it is then possible to optimize in* α*, to get better estimates.*

2.1.6.2 Confidence Interval for μ_α with σ_α^2 unknown

Suppose now that σ_α^2 is unknown. In this case, we cannot use the type of interval that we used before, so we substitute σ_α^2 with S_n^2, the *Sample Variance*

$$S_n^2 := \frac{1}{n-1}\Sigma_{j=1}^n \left(X_j - \overline{X}_n\right)^2.$$

Therefore, we look for an *Asymptotic Confidence Interval for* μ_α *at level* $1-\gamma$ of the form

$$\overline{X}_n - k\frac{S_n}{\sqrt{n}} \leq \mu_\alpha \leq \overline{X}_n + k\frac{S_n}{\sqrt{n}}.$$

We basically need to find again the value of k such that

$$P^{\frac{1}{\alpha}}\left(\overline{X}_n - k\frac{S_n}{\sqrt{n}} \leq \mu_\alpha \leq \overline{X}_n + k\frac{S_n}{\sqrt{n}}\right) \to 1-\gamma,$$

when $n \to +\infty$. Consider now the random variable

$$Z_n := \sqrt{n}\left(\overline{X}_n - \mu_\alpha\right)/S_n.$$

The distribution of this random variable is not known *a-priori*, but we can use Slutsky's Theorem.

Theorem 2.1.37 (Slutsky's Theorem)**.** *Let* $X_1, X_2, \ldots$ *and* $Y_1, Y_2, \ldots$ *be two sequence of random variables such that*

$$X_n \to X$$

in $1/\alpha$-Distribution *and*

$$Y_n \to c$$

in $1/\alpha$-Probability*, as* $n \to +\infty$*. Then*

$$X_n Y_n \to cX$$

in $1/\alpha$-Distribution*, as* $n \to +\infty$*.*

Proof. The proof is the same as in the usual Slutsky's Theorem, but replacing the usual *Convergence in Probability* and *Convergence in Distribution* with the *Convergence in* $1/\alpha$-*Probability* and *Convergence in* $1/\alpha$-*Distribution.* □

Now, Y_n converges in $\frac{1}{\alpha}$-*Distribution* to $Y \sim \mathcal{N}(0,1)$ by the *Central Limit Theorem*, while S_n^2 converges in $\frac{1}{\alpha}$-*Probability* to 1 (since $\overline{X}_n$ converges in $\frac{1}{\alpha}$-*Probability* to μ_α and $g(z) = z^2$ is a continuous function). Now,

$$Z_n = \sqrt{n}(\overline{X}_n - \mu_\alpha)/S_n = \frac{Y_n}{\sqrt{\frac{S_n^2}{\sigma_\alpha^2}}} \to \mathcal{N}(0,1)$$

in $\frac{1}{\alpha}$-*Distribution*. Therefore, proceeding as before, we have that

$$P^{\frac{1}{\alpha}}\left(\overline{X}_n - k\frac{S_n}{\sqrt{n}} \leq \mu_\alpha \leq \overline{X}_n + k\frac{S_n}{\sqrt{n}}\right) = P^{\frac{1}{\alpha}}\left(\left|\sqrt{n}\left(\overline{X}_n - \mu_\alpha\right)/S_n\right| \leq k\right) \simeq \Phi(k) - \Phi(-k).$$

If we impose again that

$$1 - \gamma = \Phi(k) - \Phi(-k)$$

we obtain that

$$k_\gamma = \Phi^{-1}\left(1 - \frac{\gamma}{2}\right)$$

and so that

$$\overline{X}_n - k_\gamma\frac{S_n}{\sqrt{n}} \leq \mu_\alpha \leq \overline{X}_n + k_\gamma\frac{S_n}{\sqrt{n}}$$

is an *Asymptotic Confidence Interval for* μ_α *at level* $1-\gamma$ when the true variance σ_α is unknown.

Remark 2.1.38. *Also here the precision of the approximation*

$$P^{\frac{1}{\alpha}}\left(\left|\sqrt{n}(\overline{X}_n - \mu_\alpha)/S_n\right| \leq k\right) \simeq \Phi(k) - \Phi(-k)$$

depends on α *and, when needed, it is then possible to optimize in* α*, to get better estimates.*

2.1.6.3 Rejection Regions and Hypothesis Testing

By means of the *Inversion Principle* (see for example Theorem 9.2.2 in [25]), we can construct *Rejection Regions* for *Hypothesis Testing* starting from *Confidence Intervals*. In particular, if we want to test the following hypothesis:

$$H_0 : \mu_\alpha = \mu_\alpha^0 \text{ vs } H_a : \mu_\alpha \neq \mu_\alpha^0,$$

we can use the results of the previous section to obtain that, if the variance σ_α^2 is known, the set

$$\left\{x : \left|\sqrt{n}(\overline{x}_n - \mu_\alpha)/\sigma_\alpha\right| > k_\gamma\right\}$$

is the *Rejection Region* for the *Most powerful test at level* γ; while, if the variance σ_α^2 is unknown,

$$\left\{x : \left|\sqrt{n}(\overline{x}_n - \mu_\alpha)/S_n\right| > k_\gamma\right\}$$

is the *Rejection Region* for the *Most powerful test at level* γ.

2.1.7 The case of the Cauchy Distribution

In this subsection, we discuss an application of the concept of $\frac{1}{\alpha}-$*Characteristic Function* to the *Cauchy Distribution*. Our aim here is to illustrate the use of the $\frac{1}{\alpha}-$*Characteristic Function* to do inference for a distribution which do not admit any finite *Momenta* for $\alpha = 1$. We first construct $\frac{1}{\alpha}$-*Momenta* for this distribution and then use the asymptotic results of the previous sections to construct *Asymptotic Confidence Interval* for the location parameter of the *Cauchy Family.*

2.1.7.1 On the $\frac{1}{\alpha}$-Momenta of order k and the Cauchy Distribution

Now, we concentrate on the case of the *Multivariate Cauchy Distribution.*

Definition 2.1.39. *We say that a random variable* X *is distributed as a* Multivariate Cauchy Distribution *if and only if its pdf takes the following form*

$$f(x; \mu, \mathbf{\Sigma}, n) = \frac{\Gamma\left(\frac{1+n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\pi^{\frac{n}{2}}\left|\mathbf{\Sigma}\right|^{\frac{1}{2}}\left[1 + (x-\mu)^T\mathbf{\Sigma}^{-1}(x-\mu)\right]^{\frac{1+n}{2}}}.$$

Here $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$, *while* $\mathbf{\Sigma}$ *is a positive definite* $n \times n$ *symmetric matrix and* $n \geq 1$ *is the dimension.*

Since this is a location scale family, we can assume $\mu = 0$ and $\Sigma = Id_{n\times n}$. Otherwise, we can just apply the affine transformation $Y := \Sigma^{1/2}(X - \mu)$ to reduce to that case. Now, we want to compute for which $\alpha > 0$ and $k, n = 1, 2, \ldots$ we can define $\frac{1}{\alpha}$ *Moment of order* k, $E^{\frac{1}{\alpha}}[X^k]$. Namely, we want to find for which values of n, k and α, we have that the quantity $\left|E^{\frac{1}{\alpha}}[X^k]\right|$ is finite. So, we do the following estimate:

$$\left|E^{\frac{1}{\alpha}}[X^k]\right| \leq C\int_{\mathbb{R}^n}\frac{|x|^k}{(1+|x|^2)^{\frac{1+n}{2\alpha}}}dx \leq C\int_1^{+\infty}\frac{\rho^{k+n-1}}{\rho^{\frac{n+1}{\alpha}}}d\rho \leq C\int_1^{+\infty}\frac{1}{\rho^{\frac{n+1}{\alpha}-n-k+1}}.$$

In this series of inequalities, the constant C can vary from step to step, but it remains independent of ρ. We have that

$$\left|E^{\frac{1}{\alpha}}[X^k]\right| < +\infty \Leftrightarrow \frac{n+1}{\alpha} - n - k + 1 > 1,$$

namely, if and only if the order of the momentum k satisfies the following condition:

$$k < \frac{n+1}{\alpha} - n.$$

In the case $n = k = 1$, we need $0 < \alpha < 1$ and in general we need $0 < \alpha < \frac{n+1}{n+k}$ to have that the $\frac{1}{\alpha}$-*Moment of order* k, $E^{\frac{1}{\alpha}}[X^k]$, is well defined.

Remark 2.1.40. *We can see that, since the number of moments k is constrained by the inequality $k < \frac{n+1}{\alpha} - n$, in the classical case $\alpha = 1$, the* Cauchy Family of Distributions *does not admit any finite* Momentum.

We can actually compute in closed form, the values of the first and second moments of the *Cauchy Distribution* for any α and n for which those moments exist. In dimension $n = 1$, to give sense to μ_α, we need $0 < \alpha < 1$, while to give sense to σ^2_α, we need $0 < \alpha < 2/3$. For simplicity, we will stay with dimension $n = 1$ and consider just the case where $\alpha = 1/p$ with p positive integer.
We start by computing the normalizing constant and we get

$$\int_{-\infty}^{+\infty} \frac{1}{\pi^p} \frac{1}{(1+x^2)^p} dx = 2\pi i R(p)$$

with

$$R(p) := \frac{1}{\pi^p} \frac{(2i)^{-2p+1}(-1)^{p-1}(2p-2)!}{[(p-1)!]^2}.$$

This is done using the *Theorem of Residues* (see for example [104]). Then $\mu_p^{2k+1} = 0$ for every k such that $k < (n+1)p - n$, because the integral of any odd function on a symmetric interval, whenever it exists, must be zero.

Remark 2.1.41. *We underline that, using the scale location invariance of the family, if X is a standard univariate Cauchy, then $Y := \mu + \sigma X$ is a Cauchy with location parameter μ and scale parameter σ. In this way, one can easily see that, for the random variable Y, $\mu^1_p = \mu$. In general, whenever it exists, $\mu_\alpha = \mu$.*

Now, we have that

$$\int_{-\infty}^{+\infty} \frac{1}{\pi^p} \frac{x^2}{(1+x^2)^p} dx = 2\pi i Q(p)$$

with

$$\begin{aligned} Q(p) \quad &:= \frac{1}{\pi^p} \frac{(2i)^{-2p+1}}{[(p-1)!]^2} \sum_{k=0}^{p-1} \frac{(p-1)!}{k!(p-1-k)!} \\ &\times [-\delta(k=0) + 2i\delta(k=1) + 2\delta(k=2)] (-1)^{p-1-k}(2i)^k(2p-k-2)!. \end{aligned}$$

Also this is done using the *Theorem of Residues*, but also by means of the differential identity

$$\frac{d^n}{dx^n}(f(x)g(x))$$

$$= \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \left[-\delta(k=0) + 2i\delta(k=1) + 2\delta(k=2)\right] (-1)^{n-k} \frac{(p+n-k-1)!}{(p-1)!},$$

with $n = p - 1$. Therefore, collecting all the previous computations, we obtain

$$Var^p[X] = \frac{2\pi i R(p)}{2\pi i Q(p)}$$

$$= \frac{\Sigma_{k=0}^{p-1} \frac{(p-1)!}{k!(p-1-k)!} \left[-\delta(k=0) + 2i\delta(k=1) + 2\delta(k=2)\right] (-1)^k (2i)^k (2p-k-2)!}{(2p-2)!}.$$

Remark 2.1.42. *We recall here that, if* c is a Pole of Order n, *then the* Residue of f around $z = c$ *can be found by the formula:*

$$\text{Res}(f, c) = \frac{1}{(n-1)!} \lim_{z \to c} \frac{d^{n-1}}{dz^{n-1}} \left((z-c)^n f(z)\right).$$

2.1.7.2 Asymptotic Results

Consider again the case $n = 1$ and $\alpha = 1/2$ (so $p = 2$). As explained in the previous section, we can compute the $\frac{1}{\alpha}$-Expected Value that is $\mu_\alpha = 0$, whenever it exists. As explained before the scale parameter μ is zero too: $\mu = \mu_\alpha = 0$. Moreover, $Var^{1/\alpha}[X] = Var^p[X] = Var^2[X] = 1$ (just by an explicit computation of the integrals or using the formula of the previous subsection). Therefore, we can apply the *Central Limit Theorem* to get that

$$\overline{X}_n \sqrt{n} \to \mathcal{N}(0, 1)$$

in $2-$*Distribution*. This enables us to do some inference as presented in Section 2.1.6 and construct, for this distribution, both *Confidence Intervals* for $\mu = \mu_\alpha$ and *Rejection Regions* for *Hypothesis Testing* for $\mu = \mu_\alpha$ again as presented in Section 2.1.6. Again, by using the scale-location invariance of the family, if X is a *Standard Cauchy Distribution*, then $Y := \mu + \sigma X$ is a *Cauchy Distribution* with location parameter μ and scale parameter σ. In this way, we can easily see that, for the random variable Y, $\mu^\alpha = \mu_p^1 = \mu$ and so the confidence interval for μ_p^1 becomes an *Asymptotic Confidence Interval for* μ *at level* $1 - \gamma$ of the form:

$$\overline{X}_n - k_\gamma \frac{1}{\sqrt{n}} \le \mu \le \overline{X}_n + k_\gamma \frac{1}{\sqrt{n}}.$$

Similar results can be deduced in the case of unknown variance and for different values of α.

Remark 2.1.43. *In [89] the author employs a different and elegant method to estimate the parameters of the* Univariate Cauchy Distribution. *He takes advantage of the equivariance under* Möbius Transformations *of the* Cauchy Family*, but he explains that, due to this fact, his strategy unlikely will extend to other family of distributions. Differently from [89], our method does not rely on any particular algebraic/geometric structure of a single family (like the equivariance mentioned above) and so it works for any family of distribution. Moreover, our method work in any dimensions and not just in the univariate case as for [89].*

Remark 2.1.44. *We discussed here several results about* $\frac{1}{\alpha}$-Characteristic Functions *and their application, but we believe that lots of other results can be proved and that this tool can be very helpful in a very general framework.*

2.2 The Monge-Ampère Equation in Transformation Theory and an Application to $\frac{1}{\alpha}$-Probabilities

In this section, we give an application of the theory of Monge-Ampère Equations to *Transformation Theory* in Probability. The motivation behind this section is to show that the transformation from a pdf to its $\frac{1}{\alpha}$-counterpart is transferred also at the level of the random variables.

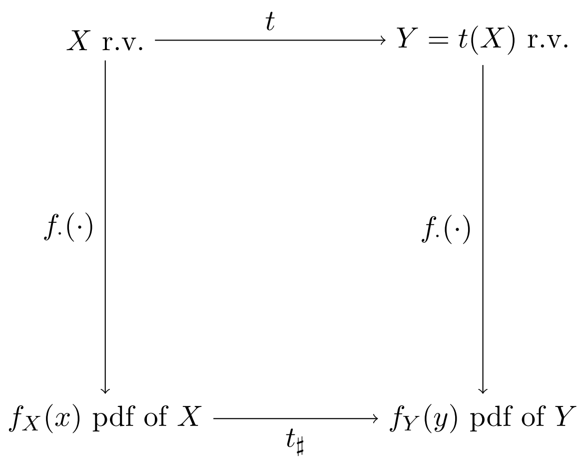
We treat explicitly the cases of the *Multivariate Normal Distribution*, *Multivariate Exponential Distribution* and *Cauchy Distribution*. Moreover, we prove some rigidity theorems on the possible transformations which send a pdf to its $\frac{1}{\alpha}$-counterpart. In the general case, it is not possible to construct explicitly the transformation between the random variables, despite it is always possible to reduce to quadrature the transformation between the pdfs. We conclude the section by setting our theorems in the context of the Monge-Ampère equations and the Optimal Transportation theory and by giving some numerics to illustrate the motion of mass, while transforming a pdf to its $\frac{1}{\alpha}$-counterpart.

2.2.1 Introduction and Motivation

An important issue in probability and statistics is the one of *Transformation Theory*. The main problems addressed by *Transformation Theory* are mainly two:

- (*Direct Problem*) Suppose you have a random variable X with its pdf $f_X(x)$ and another random variable Y which *explicitly* depends on X through a transformation t. Then you can write $Y = t(X)$. What is the pdf $f_Y(y)$ of the random variable Y? How does it depend on the transformation t?

- (*Inverse Problem*) Suppose now that you have a rule which assigns to the pdf $f_X(x)$ of X another pdf $f_Y(y)$. Is there a transformation t, such that $Y = t(X)$ is a random variable with pdf $f_Y(y)$? Can you construct *explicitly* this transformation t?



The *Direct Problem* is very well understood and in fact we have the very general and very well known theorem (see for example [25]).

Theorem 2.2.1. *[Change of Variables Formula] Let X be a random variable with pdf $f_X(x)$ and let $Y = t(X)$. Here t is an invertible transformation from the range of X to the range of Y. Define the sets $\mathcal{X}$ and $\mathcal{Y}$ in the following way:*

$$\mathcal{X} := \{x : f_X(x) > 0\} \text{ and } \mathcal{Y} := \{y : y = t(x) \text{ for some } x \text{ in } \mathcal{X}\}.$$

Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $t^{-1}(y)$ has continuous derivatives on $\mathcal{Y}$. Then the pdf of Y is given by the following formula:

$$f_Y(y) = f_X(t^{-1}(y))|det(J(t^{-1}(y))| \quad y \in \mathcal{Y} \tag{2.10}$$

and 0 otherwise. Here Jt^{-1} is the Jacobian Matrix *of the transformation t^{-1}, inverse of t.*

The *Inverse Problem* is much more complicated and leads to a Partial Differential Equation very hard to solve, because it is *Fully Nonlinear*. This equation goes under the name of *Monge-Ampère Equation* and we can easily deduce it as follows. As hypothesis of the *Inverse Problem*, we know both $f_X(x)$ and $f_Y(y)$. On the set $\mathcal{Y}$, $f_X(t^{-1}(y)) \neq 0$ by definition, so dividing (2.10) on both sides by $f_X(t^{-1}(y))$ we get:

$$|det(Jt^{-1}(y))| = \frac{f_Y(y)}{f_X(t^{-1}(y))} \quad y \in \mathcal{Y}.$$

Now, since the transformation t is invertible and smooth, then both $|det(Jt(y)| \neq 0$ and $|det(Jt^{-1}(y)| \neq 0$. If we *assume that t is regular enough*, then $det(Jt(y))$ is a regular function of y, so if it is never zero, it must keep the same sign for every $y \in \mathbb{R}^n$. Therefore, $det(Jt(y))$ is always either positive or negative and this leads to the *Monge-Ampère Equation*:

$$det(Jt^{-1}(y)) = F(y) \quad y \in \mathcal{Y} \tag{2.11}$$

with $y \in \mathbb{R}^n$, $F(y) = +\frac{f_Y(y)}{f_X(t^{-1}(y))}$ or $F(y) = -\frac{f_Y(y)}{f_X(t^{-1}(y))}$ if, respectively, $det(Jt(y)) > 0$ or $det(Jt(y)) < 0$. A simple application of the theory of *Monge-Ampère Equations* give us the following results.

Theorem 2.2.2. *[Monge-Ampère Equation] Suppose that $f_X(x) : \mathbb{R}^n \to \mathbb{R}$ and $f_Y(y) : \mathbb{R}^n \to \mathbb{R}$ are the pdfs of two real random variables such that $t_\sharp f_X = f_Y$ with $t_\sharp$ explicit. Then, the transformation t such that $Y = t(X)$, and which satisfies $t_\sharp f_X = f_Y$, must be a solution of the following differential equations:*

- *$n = 1$: t must satisfy the* Semilinear Ordinary Differential Equation

$$\frac{d}{dy}t^{-1}(y) = \frac{f_Y(y)}{f_X(t^{-1}(y))} \quad y \in \mathcal{Y};$$

- *$n \geq 2$: t must satisfy the* Fully Nonlinear Monge-Ampère Equation

$$det(Jt^{-1}(y)) = \frac{f_Y(y)}{f_X(t^{-1}(y))} \quad y \in \mathcal{Y}.$$

Here, $\mathcal{Y}$ is defined as follows:

$$\mathcal{Y} := \{y : y = t(x) \text{ for some } x \text{ in } \mathcal{X}\}$$

and

$$\mathcal{X} := \{x : f_X(x) > 0\}.$$

Remark 2.2.3. *From the theorem, it is pretty clear that the lower the dimension n is, the simpler the problem becomes. This is true, because, in low dimension, it is simpler to solve the differential equation that the transformation t has to satisfy.*

Now, we apply this theorem in the context of $\frac{1}{\alpha}$-*Characteristic Functions*. The $\frac{1}{\alpha}$-*Characteristic Function* is a tool introduced in Section 2.1 as a natural generalization of the classical *Characteristic Function*. The main novelty of this tool is that it permits to extend classical theorems, such as the *Law of Large Numbers* (LLN) and the *Central Limit Theorem* (CLT) to basically every distribution, upon the correct choice of a free parameter α. This allows us to do some *Asymptotic Inference* with distributions which do not have any finite classical moments, such as finite *mean* or finite *variance* (for example). Using Theorem 2.2.2, we can compute explicitly some of the transformations which send a pdf to its $\frac{1}{\alpha}$-counterpart.

Corollary 2.2.4. *[Explicit Transformations] Suppose that X is a random variable whose pdf is $f_X(x)$. Suppose also that $f_Y(y) = f_X^{\frac{1}{\alpha}}(y)$, where*

$$f_X^{\frac{1}{\alpha}}(x) := \frac{f_X(x)^{1/\alpha}}{\int_{\mathbb{R}^n} f_X(x)^{\frac{1}{\alpha}} dx}$$

is the $\frac{1}{\alpha}$-counterpart of $f_X(x)$. Then, we can construct an explicit *invertible and smooth map t such that $Y = t(X)$ and the pdf of Y is $f_Y(y)$ in the following cases:*

- *if X is distributed as a* Multivariate Normal Distribution *$X \sim MVN(0, \Sigma)$, then $u(x) := t(x) = (x_1\sqrt{\alpha}, \dots, x_n\sqrt{\alpha}) = \sqrt{\alpha}x$.*

- *if X is distributed as a* Multivariate Exponential Distribution *$X \sim MVE(\Lambda)$, with $\Lambda = (\lambda_1, \dots, \lambda_n)$, then $u(x) := t(x) = \alpha(x_1, \dots, x_n) = \alpha x$.*

- *if X is distributed as a* Cauchy Distribution *$X \sim Cauchy(0)$ and $\alpha = \frac{1}{2}$, then $x = t^{-1}(y) = tan\left(\frac{y}{1+y^2} + arctan(y)\right)$.*

Remark 2.2.5. *Similar constructions work in more general cases, like for example the case of the* Cauchy Distribution *with a different α (the proofs follow the same lines). Since the main purpose of this section is to point out certain important facts of the theory, we do not treat as many cases as we can and we leave some further applications to a future work.*

Another question to which we want to answer is the following. In Corollary 2.2.4, we basically treated the problem distribution by distribution. The reason is that there is not an explicit formula for the solutions of the *Fully Nonlinear Monge Ampère Equation* and, therefore, we do not have a general explicit formula that gives the transformation between the random variables once we know the transformation between the pdfs. Now, a natural question is: when we pass from a pdf $f_X(x)$ of a random variable X to its $\frac{1}{\alpha}$-counterpart, is there a transformation t such that the random variable $Y = t(X)$ has, as its own pdf, the density $f_X^{\frac{1}{\alpha}}(y)$ with $t(X) = bX^a$? The answer is given by the following theorem.

Theorem 2.2.6. *[Rigidity Theorem] Fix the dimension $n = 1$. Suppose that $f_X(x)$ is the pdf of a random variable X and $f_X^{\frac{1}{\alpha}}(y)$ is its $\frac{1}{\alpha}$-counterpart. Suppose moreover that there is a transformation $t : \mathbb{R} \to \mathbb{R}$ such that the random variable $Y = t(X)$ has, as its own pdf, the density $f_Y(y) = f_X^{\frac{1}{\alpha}}(y)$ with $t(X) = bX^a$. Then, the following facts are true.*

1. *Both a and b cannot be taken universal and they depend on $f_X(x)$, the particular pdf of the random variable X, since, otherwise, they are trivial: $a = b = \alpha = 1$.*

2. *Suppose $a = 1$ and $\alpha \neq 1$, then the pdf $f_X(x)$ is of the form*

$$f_X(x) = d^{\frac{\alpha}{1-\alpha}} e^{-\frac{K_2(\alpha)}{p+1} x^{p+1}},$$

   where d, p and $K_2(\alpha)$ are real constants which depend on α.

Remark 2.2.7. *Theorem 2.2.6 confirms that a polynomial transformation at the level of the pdfs do not correspond in general to a polynomial transformation at the level of the random variables.*

Remark 2.2.8. *From Theorem 2.2.6, it emerges a family of distributions that one can call* Uniform-Exponential Family*. It can be indexed not just by α, but also by other parameters like for example p. The second part of Theorem 2.2.6 can be viewed as a characterization of the distributions of the* Uniform-Exponential Family *with pdfs of the form $f_X(x) = d^{\frac{\alpha}{1-\alpha}} e^{-\frac{K_2(\alpha)}{p+1} x^{p+1}}$.*

Remark 2.2.9. *We believe that the possibility of building these explicit transformations, between a random variable and its $\frac{1}{\alpha}$-counterpart, can be very useful in the cases in which the inference is non-trivial because of the non-suitable decay or regularity of the original pdf. In fact, in these not easy cases, one can explicitly pass to the $\frac{1}{\alpha}$-pdf, do inference there and then transform back the results.*

The remaining part of the section is organized as follows. In Subsection 2.2.2, we give the proofs of Theorem 2.2.2 and of its Corollary 2.2.4 and also of Theorem 2.2.6. In Subsection 2.2.3, we collect and explain some results about the *Monge-Ampère Equation.*

In Subsection 2.2.4, we restate our results in the context of *Optimal Transportation Theory*, while in Subsection 2.2.5, we do some numerics and give some comments on how the point of view of *Transportation Theory* explains the movement of mass from one pdf to its $\frac{1}{\alpha}$-counterpart.

2.2.2 Proofs

In this subsection, we give the proofs of our main results. For what concerns Theorem 2.2.1 and Theorem 2.2.2, we do not enter into the details of the proofs either because they are trivial, or because they have been already given in the Introduction and they are well known. Instead, we give the complete proof of Corollary 2.2.4 and Theorem 2.2.6.

2.2.2.1 Proof of Corollary 2.2.4

The strategy here is to find a particular solution of the *Fully Nonlinear Monge-Ampère Equation*

$$det(Jt^{-1}(y)) = \frac{f_Y(y)}{f_X(t^{-1}(y))} \quad y \in \mathcal{Y}$$

in the cases under investigation. We first treat the case of the *Multivariate Normal Distribution*. Suppose $n = 2$ and $X \sim MVN(0, Id)$, so $f_X(x, y) = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$. From this, we can construct $f^{\frac{1}{\alpha}}(x, y)$ and get

$$f^{\frac{1}{\alpha}}(x, y) = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2\alpha}}\alpha^{-1},$$

therefore we obtain

$$\frac{f_X(x, y)}{f^{\frac{1}{\alpha}}(u)} = \frac{\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}}{\frac{1}{2\pi}e^{-\frac{u_1^2+u_2^2}{2\alpha}}\alpha^{-1}} = \alpha e^{\frac{x^2+y^2}{2} - \frac{u_1^2+u_2^2}{2\alpha}},$$

where $u := t^{-1}$. We arrive at the equation

$$\frac{\partial u_2}{\partial y}\frac{\partial u_1}{\partial x} - \frac{\partial u_2}{\partial x}\frac{\partial u_1}{\partial y} = \alpha e^{\frac{x^2+y^2}{2} - \frac{u_1^2+u_2^2}{2\alpha}}.$$

If we choose $u_1 = \alpha^{-1/2}x$ and $u_2 = \alpha^{-1/2}y$, the equation is satisfied and so $u(x, y) = t^{-1}(x, y) = \alpha^{-1/2}(x, y)$. In the case of general dimension n, we notice that in the formula of $det(Ju(y))$, there exists just one term with all the $\frac{\partial u_i}{\partial y_i}$ for every $i = 1, \ldots, n$. Therefore, if we choose again $u_i(y) = u_i(y_i)$ for every $i = 1, \ldots, n$, then

$$det(Ju(y)) = \Pi_{i=1}^n \frac{\partial u_i}{\partial y_i}$$

and so we have just to solve

$$\Pi_{i=1}^{n}\frac{\partial u_i}{\partial y_i} = \alpha^{n/2}e^{\frac{\Sigma_{i=1}^{n}y_i^2}{2}-\frac{\Sigma_{i=1}^{n}u_i^2}{2\alpha}},$$

which, again, admits the solution: $u_i = \alpha^{+1/2}y_i$ and so $u(y) = t^{-1}(y) = \alpha^{+1/2}y$. The case of the *Multivariate Normal Distribution* is solved.

Now, we treat the case of the *Multivariate Exponential Distribution*. Suppose $X \sim MVE(\lambda)$, so $f_X(y) = \lambda^n e^{-\lambda\Sigma_{i=1}^{n}y_i}$. From this we can construct $f^{\frac{1}{\alpha}}(y)$ and get

$$f^{\frac{1}{\alpha}}(y) = \frac{\lambda^{n/\alpha}e^{-\lambda\alpha^{-1}\Sigma_{i=1}^{n}y_i}}{\lambda^{n/\alpha}\left(\frac{\lambda}{\alpha}\right)^{-n}}.$$

Therefore, we obtain

$$\frac{f_X(y)}{f^{\frac{1}{\alpha}}(u)} = \frac{\lambda^n e^{-\lambda\Sigma_{i=1}^{n}y_i}}{\frac{\lambda^{n/\alpha}e^{-\lambda\alpha^{-1}\Sigma_{i=1}^{n}u_i}}{\lambda^{n/\alpha}\left(\frac{\lambda}{\alpha}\right)^{-n}}} = \alpha^n e^{-\lambda\Sigma_{i=1}^{n}y_i-\alpha^{-1}u_i(y)}$$

and so, defining again $u := t^{-1}$, the equation becomes

$$det(Ju(y)) = \alpha^n e^{-\lambda\Sigma_{i=1}^{n}(y_i-\alpha^{-1}u_i(y))}.$$

Again, we notice that in the formula of $det(Ju(y))$, there exists just one term of the sum which involves all the $\frac{\partial u_i}{\partial y_i}$ for every $i = 1, \dots, n$. So, if we choose again $u_i(x) = u_i(x_i)$ for every $i = 1, \dots, n$, then

$$det(Ju(y)) = \alpha^n e^{-\lambda\Sigma_{i=1}^{n}(y_i-\alpha^{-1}u_i(y))}$$

and so we have just to solve

$$\Pi_{i=1}^{n}\frac{\partial u_i}{\partial y_i} = \alpha^n e^{-\lambda\Sigma_{i=1}^{n}y_i-\alpha^{-1}u_i(y)},$$

which admits the solution $u_i = \alpha y_i$ and so $u(y) = t^{-1}(y) = \alpha y$. This completes the proof in the case of the *Multivariate Exponential Distribution*.

Now, it remains to prove the case of the *Cauchy Distribution*, with $n = 1$, $\alpha = \frac{1}{2}$. In this case, X is such that its pdf takes the form $f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$. We can again, from this, construct $f_X^{\frac{1}{\alpha}}(x)$ and get

$$f_X^{\frac{1}{\alpha}}(x) = \frac{\pi^{-1/\alpha}(1+x^2)^{-1/\alpha}}{\int_{-\infty}^{+\infty}[f_X(x)]^{1/\alpha}dx} = \frac{2}{\pi}(1+x^2)^{-2}.$$

Therefore, we get

$$\frac{f_X(x)}{f^{\frac{1}{\alpha}}(u)} = \frac{\pi^{-1}(1+x^2)}{\pi^{-1}2(1+u(x)^2)^{-2}} = \frac{1}{2}\frac{(1+u(x)^2)^2}{1+x^2}$$

and so we obtain the equation

$$\frac{du(x)}{dx} = \frac{1}{2}\frac{(1+u(x)^2)^2}{1+x^2}.$$

By separation of variables, we then get

$$\frac{du}{(1+u(x)^2)^2} = \frac{1}{2}\frac{dx}{1+x^2}$$

that, with the condition $u(0) = 0$, integrates to:

$$\frac{1}{2}\left(\frac{u}{1+u^2} + arctan(u)\right) = \frac{1}{2}arctan(x.)$$

Therefore, by inverting $arctan$, we get

$$x = t(y) = tan\left(\frac{y}{1+y^2} + arctan(y)\right).$$

Now, if you compute the first derivative of this transformation, we get

$$\frac{dy}{dx} = \frac{2\sec\left(\frac{x}{1+x^2} + arctan(x)\right)}{(1+x^2)^2}.$$

This implies that $\frac{dy}{dx} > 0$ for every x, and so it is invertible when it exists. This concludes the proof of the corollary.

Remark 2.2.10. *We think that the construction of explicit solutions with the technique used above can be of independent interest. See [26] for more results on entire solutions of Monge-Amp ére equations.*

2.2.2.2 Proof of Theorem 2.2.6

In this subsection, we prove Theorem 2.2.6. Suppose that we have two random variables X and Y such that $Y = bX^a$ with X a random variable with cdf $F_X(x)$, then we have:

$$F_Y(y) = P(Y \leq y) = P(bX^a \leq y) = P\left(X \leq \frac{1}{b}y^{\frac{1}{a}}\right) = F_X\left(\frac{1}{b}y^{\frac{1}{a}}\right).$$

On the other side, by hypothesis, $f_Y(y) = \frac{1}{K(\alpha)}[f_X(y)]^{\frac{1}{\alpha}}$, where $K(\alpha) = \int_{\mathbb{R}}[f_X(x)]^{\frac{1}{\alpha}}dx$. Therefore, we obtain the functional identity:

$$\left[\frac{K(\alpha)}{ab} y^{\frac{1}{a}-1} f\left(\frac{1}{b} y^{\frac{1}{a}}\right)\right]^{\alpha} = f(y),$$

with $f(x) := f_X(x)$. Here, we are implicitly using the hypothesis $y > 0$, but a similar treatment works for $y < 0$. Changing variables with $z := \frac{1}{b} y^{\frac{1}{a}}$, we obtain

$$\left[\frac{K(\alpha)}{ab}\right]^{\alpha} z^{\alpha(1-a)} [f(z)]^{\alpha} = f(bz^a),$$

for every $z \in \mathbb{R}$. Since we want this to be true for every f, then we just test this identity over different distributions and show that $Y = bX^a$ implies $a = b = \alpha = 1$.
Suppose $f_X(x) = 1/c$ for $x \in [0, c]$, namely $X \sim Unif([0, c])$. Then

$$\left[\frac{K(\alpha)}{ab^a}\right]^{\alpha} z^{\alpha(1-a)} [f(z)]^{\alpha} = f(b^a z^a)$$

becomes

$$\left[\frac{K(\alpha)}{ab^a}\right]^{\alpha} z^{\alpha(1-a)} = c^{\alpha-1},$$

for every $z \in [0, c]$. Now, for $z = c$, we get

$$\left[\frac{K(\alpha)}{ab^a}\right]^{\alpha} = c^{a\alpha-1}.$$

For $z = c/2$, we get instead

$$\left[\frac{K(\alpha)}{ab^a}\right]^{\alpha} = 2^{\alpha(a-1)} c^{a\alpha-1}.$$

Putting everything together, we then get

$$\left[\frac{K(\alpha)}{ab^a}\right]^{\alpha} = 2^{\alpha(a-1)} \left[\frac{K(\alpha)}{ab^a}\right]^{\alpha}$$

and so $a = 1$, since $K(\alpha) \neq 0$. With this constraint on a, we are reduced to

$$\left[\frac{K(\alpha)}{b}\right]^{\alpha} [f(z)]^{\alpha} = f(bz).$$

Now, we consider $f_X(x) = e^{-x}$, namely $X \sim exp(1)$ and this implies

$$\left[\frac{K(\alpha)}{b}\right]^{\alpha} e^{-z\alpha} = e^{-bz}.$$

By taking the logarithm to both sides (they are positive), we obtain

$$z(\alpha - b) = \alpha \log\left(\frac{K(\alpha)}{b}\right).$$

Since the equality must hold for every z, then $\alpha - b = 0$. This implies that $\alpha = b = K(\alpha)$ and so our identity reduces to

$$[f(z)]^{\alpha} = f(\alpha z),$$

for every $z \in \mathbb{R}$. If we now test this identity with the standard normal distribution, we obtain:

$$\left(\frac{1}{\sqrt{2\pi}}\right)^{\alpha} e^{-\frac{\alpha}{2}z^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\alpha^2 z^2}$$

and this implies $\alpha = 1$ and so $a = b = 1$, namely the trivial case $Y = X$ and $\alpha = 1$.
Now, we pass to the second part of the theorem. Suppose $t(x) = cx$, then the the *Fully Nonlinear Monge-Ampère Equation*

$$det(Jt^{-1}(y)) = \frac{f_Y(y)}{f_X(t^{-1}(y))} \quad y \in \mathcal{Y}$$

becomes

$$\frac{c}{n(\alpha)}[f(cy)]^{\frac{1}{\alpha}} = f(y),$$

where $n(\alpha) = \int_{\mathbb{R}}[f(x)]^{\frac{1}{\alpha}}dx$. To make the notation lighter define $d := \frac{c}{n(\alpha)}$. Recall that c is actually a function of α, so $c = c(\alpha)$. Since both sides of $\frac{c}{n(\alpha)}[f(cy)]^{\frac{1}{\alpha}} = f(y)$ are positive, we can take the logarithm to both sides and, after defining $w(y) := \log(f(y))$, we obtain

$$w(cy) = \alpha w(y) - \alpha \log(d).$$

Now, we take a derivative with respect to y to both the sides (here $w'(z) := \frac{d}{dz}w(z)$) and obtain:

$$cw'(cy) = \alpha w'(y),$$

since $\alpha \log(d)$ is constant in y. Now, we define $h(z) := w'(z)$ and we differentiate two times with respect to α (note that in the following $\dot{c} := \frac{d}{d\alpha}c(\alpha)$) both sides of $ch(cy) = \alpha h(y)$, the identity obtained. This gives:

$$\ddot{c}h(cy) + (\dot{c})^2 y h'(cy) + cy\ddot{c}h'(cy) + y^2 c(\dot{c})^2 h''(cy) + y(\dot{c})^2 h'(cy) = 0.$$

Changing variables to $z := yc$, we obtain the identity:

$$\ddot{c}ch(z) + [2(\dot{c})^2 + c\ddot{c}]zh'(z) + z^2(\dot{c})^2 h''(z) = 0.$$

This is a second order linear ODE in the unknown $h(y)$ with smooth coefficients, so it admits a two parameter family of solutions. Now, we have to distinguish between two different cases.

In the case $-1 \neq -\frac{c\ddot{c}}{(\dot{c})^2}$, the general solution is given by the following family

$$h(z) = K_1 z^{-1} + K_2 z^{-\frac{c\ddot{c}}{(\dot{c})^2}}.$$

This can be verified for example by using the *ansatz* $h(z) = z^p$ and finding that $p = -1$ or $p = -\frac{c\ddot{c}}{(\dot{c})^2}$. Since $h(z) = w'(z)$, then $w(z) = K_1 \log(z) + \frac{K_2}{p+1} z^{p+1} + K_3$ with $p = -\frac{c\ddot{c}}{(\dot{c})^2}$ and hence, since $w(z) = \log(f(z))$, we have:

$$f(z) = e^{K_3} z^{K_1} e^{\frac{K_2}{p+1} z^{p+1}},$$

for some real constants K_1, K_2 and K_3. We now want to find the constants such that the condition

$$w(cy) = \alpha w(y) - \alpha \log(d)$$

is satisfied. Plugging our function $w(z)$ inside this equation, we get

$$K_1 \log(c) + K_1 \log(y) + \frac{K_2}{p+1}(cy)^{p+1} + K_3 = \alpha K_1 \log(y) + \alpha \frac{K_2}{p+1} y^{p+1} + \alpha K_3 - \alpha \log(d).$$

Collecting everything term by term, we get the following identity

$$[K_1(1-\alpha)] \log(y) + y^{p+1} K_2 \left[\frac{c^{p+1}}{p+1} - \frac{\alpha}{p+1} \right] + [K_1 \log c + K_3(1-\alpha) + \alpha \log d] = 0,$$

for every $y \in \mathbb{R}$. This is a functional identity and so it implies that the coefficients of the left hand side must be all zero. We suppose that $\alpha \neq 1$, otherwise we come back to the trivial case. This implies right away that $K_1 = 0$. If $K_2 = 0$ we get $f(y) = e^{K_3}$ with $K_3 = \frac{\alpha \log d}{\alpha - 1}$ so $f(x) = d^{\frac{\alpha}{\alpha-1}}$ is the pdf of a *Uniform Distribution*. If $K_2 \neq 0$, then

$c(\alpha) = \alpha^{\frac{1}{p+1}}$ and $K_3(1-\alpha) + \alpha \log d = 0$, so $K_3 = \frac{\alpha \log d}{\alpha - 1}$ and

$$f(z) = d^{\frac{\alpha}{\alpha-1}} e^{-\frac{K_2(\alpha)}{p+1} z^{p+1}},$$

with $K_2(\alpha)$ determined through the normalizing condition and the constant $c(\alpha)$ implicitly determined by $c(\alpha) = \alpha^{\frac{1}{p+1}}$ with $p = -\frac{c\ddot{c}}{(\dot{c})^2}$, determined through the condition $\frac{c}{n(\alpha)}[f(cy)]^{\frac{1}{\alpha}} = f(y)$. We note that the condition $c(\alpha) = \alpha^{\frac{1}{p+1}}$ with $p = -\frac{c\ddot{c}}{(\dot{c})^2}$ is not void, since it is satisfied at least by $c(\alpha) = \alpha$ and $c(\alpha) = \alpha^{1/2}$ (but also for any $c(\alpha) = \alpha^N$, $N \in \mathbb{R}$), in accordance with Corollary 2.2.4 and just by direct substitution).
In the case $-1 = -\frac{c\ddot{c}}{(\dot{c})^2}$, which happens if and only if $c(\alpha) = L_2 e^{L_1 \alpha}$ ($L_1, L_2 \in \mathbb{R}$ are constants of integration), the general solution is given by the following family

$$h(z) = K_1 \frac{1}{z} + K_2 \frac{log(z)}{z}.$$

This can be verified for example by direct computation. Since $h(z) = w'(z)$, then

$$w(z) = K_1 \log(z) + \frac{K_2}{2} \log^2(z) + K_3$$

and hence, since $w(z) = \log(f(z))$, we have:

$$f(z) = e^{K_3} z^{K_1} z^{\log\left(\frac{K_2}{2}\right)},$$

for some real constants K_1, K_2 and K_3. We now want to find the constants such that the condition

$$w(cy) = \alpha w(y) - \alpha \log(d)$$

is satisfied. Plugging our function $w(z)$ inside this equation, we get

$$K_1 \log(c) + K_1 \log(y) + \frac{K_2}{2} \left(\log(c) + \log(y)\right)^2 + K_3$$

$$= \alpha K_1 \log(y) + \alpha \frac{K_2}{2} \log^2(y) + \alpha K_3 - \alpha \log(d).$$

Collecting everything term by term, we get the following identity

$$\left(K_1 \log(c) + \frac{K_2}{2} \log^2(c) - \alpha K_3 + \alpha \log(d)\right) +$$
$$\log(y)\left(K_2 \log(c) + (1-\alpha) K_1\right) + \log^2(y) \left(\frac{K_2}{2} - \alpha \frac{K_2}{2}\right) = 0,$$

for every $y \in \mathbb{R}$. This is a functional identity and so it implies that the coefficients of the left hand side must be all zero. The condition $\left(\frac{K_2}{2} - \alpha\frac{K_2}{2}\right) = 0$ implies $K_2 = 0$ and so this, with $(K_2 \log(c) + (1-\alpha)K_1) = 0$ implies that $K_1 = 0$. Using

$$\left(K_1 \log(c) + \frac{K_2}{2} \log^2(c) - \alpha K_3 + \alpha \log(d)\right) = 0,$$

we get $K_3 = \log(d)$ and so $f(z) = d = \frac{c}{n(\alpha)}$. Now, using the condition

$$\frac{c}{n(\alpha)}[f(cy)]^{\frac{1}{\alpha}} = f(y),$$

we get: $d \times d^{\frac{1}{\alpha}} = d$ and so, since $d \neq 0$ by hypothesis,

$$\frac{c}{n(\alpha)} = d = 1.$$

Now, $n(\alpha) = c = L_1 e^{L_2 \alpha}$, but also

$$n(\alpha) = \int_{\mathbb{R}} [f(x)]^{\frac{1}{\alpha}} dx = \left(\frac{c}{n(\alpha)}\right)^{\frac{1}{\alpha}} \times \left(\frac{c}{n(\alpha)}\right)^{-1} = \left(\frac{c}{n(\alpha)}\right)^{\frac{1}{\alpha}-1}$$

and so $c = n(\alpha) = c^{1-\alpha}$, that implies $\alpha = 0$, which is impossible, or $c = 1$, which is the trivial case of the identity transformation. This concludes the proof of the theorem.

Example 2.2.11. *Consider the case $p = 0$, this implies $-\frac{c\ddot{c}}{(\dot{c})^2} = 0$ and since $c \neq 0$ (otherwise we go back to the trivial case), $c(\alpha) = a_1\alpha + a_2$ for some real constants a_1 and a_2, therefore*

$$f(z) = \left[\frac{a_1\alpha + a_2}{n(\alpha)}\right]^{\frac{\alpha}{\alpha-1}} e^{-K_2(\alpha)z}.$$

Due to the normalizing condition, we have

$$f(z) = K_2(\alpha) e^{-K_2(\alpha)z}$$

with $K_2(\alpha) = \left[\frac{a_1\alpha+a_2}{n(\alpha)}\right]^{\frac{\alpha}{\alpha-1}}$. Since we can choose $a_1 = 1$ and $a_2 = 0$, we also have that $n(\alpha) = \int_0^{+\infty} K_2(\alpha)e^{-K_2(\alpha)z}dz = \alpha$. This implies that $K_2(\alpha) = \alpha$ and so that $f(z) = e^{-z}$, namely the pdf of the Exponential Distribution*. This confirms the result of Corollary 2.2.4. If we ask $p = 1$, we get $c(\alpha) = \sqrt{a_1\alpha + a_2}$ for some real constants a_1 and a_2 and we recover the Normal Distribution, again as in Corollary 2.2.4.*

Remark 2.2.12. *This theorem basically says that if there is a transformation t of the* *form* $t(X) = bX^a$, *then both* a *and* b *depend on the pdf* f *through the equation*

$$\left[\frac{K(\alpha)}{ab}\right]^{\alpha} z^{\alpha(1-a)}[f(z)]^{\alpha} = f(bz^a).$$

2.2.3 The Monge-Ampère Equation

The *Monge-Ampère Equation* is a very well know equation in geometry and mathematical physics. It arises naturally in several problems in *Riemannian Geometry*, *Conformal Geometry*, *Complex-Kähler Geometry* and *CR geometry.* The *Monge-Ampère Equations* are related to the *Monge-Kantorovich Optimal Mass Transportation Problem*, when the *cost functional* therein is given by the Euclidean distance (see Section 2.2.4 and for example [52]).

In this subsection, we collect some basic results of the theory, just for illustration purposes. We start by giving the general definition of *Monge-Ampère Equation* in dimension $n = 2$.

Definition 2.2.13. *Given two independent variables* x *and* y, *and one dependent variable* u, *the general* Monge-Ampère Equation *is of the form*

$$L[u] = A(u_{xx}u_{yy} - u_{xy}^2) + Bu_{xx} + 2Cu_{xy} + Du_{yy} + E = 0$$

where A, B, C, D *and* E *are functions depending on the variables* x, y, ϕ, u_x, *and* u_y *only.*

Remark 2.2.14. *In the previous sections, we used mainly the case where* $u = (u_1, u_2)$ *with* $u_1 = u_x$ *and* $u_2 = u_y$, *and coefficients* $A = 1$, $B = C = D = 0$ *and* $E = -\frac{f(y)}{f^{\frac{1}{\alpha}}(u(y))}$.

Theorem 2.2.15 (Rellich's Theorem)**.** *Let* Ω *be a bounded domain in* $\mathbb{R}^3$, *and suppose that, on* Ω, A, B, C, D, *and* E *are continuous functions of* x *and* y *only. Consider the Dirichlet problem:*

$$\begin{cases} L[u] = 0, & \text{on } \Omega \\ u|_{\partial\Omega} = g(x, y). \end{cases}$$

Here g *is a smooth function. If*

$$BD - C^2 - AE > 0,$$

then the Dirichlet problem has at most two solutions u *which assume the same boundary value on* $\partial\Omega$.

Remark 2.2.16. *This theorem cannot be generally applied to solve our main theorems, because in our cases* $E = -\frac{f(y)}{f^{\frac{1}{\alpha}}(u(y))}$ *and so it depends on the unknown function* $u(y)$.

If the source function E does not depend on the unknown function, but possesses certain symmetries, it is sometimes possible to construct solutions of the *MongeAmpère equation*, provided we can solve a *Quasilinear ODE.*

Remark 2.2.17. *Fix the dimension* $n = 2$ *and consider a source of the form* $f(x, y) = F(ax^2 + bxy + cy^2 + kx + sy)$, *for some real constants* a, b, c, k, s. *Then the* Monge-Ampère equation *is reduced to*

$$\frac{\partial^2 u}{\partial x \partial x}\frac{\partial^2 u}{\partial y \partial y} - \frac{\partial^2 u}{\partial x \partial y}\frac{\partial^2 u}{\partial y \partial x} = F(ax^2 + bxy + cy^2 + kx + sy).$$

At this point, we can look for solutions of the form $u(x, y) = W(z)$ *with* $z = ax^2 + bxy + cy^2 + kx + sy$ *and* W *solving the following quasilinear ODE:*

$$2[(4ac - b^2)z + as^2 + ck^2 - bks]W'(z)W''(z) + (4ac - b^2)[W'(z)]^2 + F(z) = 0.$$

With the substitution $w(z) := [W'(z)]^2$ *and, for example, with the choice of parameters* $b = k = s = 0$ *and* $a = c = 1$ *(spherical case) one is lead to the following linear ODE:*

$$8w'(z) + 4w(z) + F(z) = 0$$

that can be solved by the method of Variation of Constants. *For several other cases that can be of interest and where it is possible to find an explicit solution of the* Monge-Ampère equation, *we refer to the handbook [96].*

2.2.4 Connection to the problem of Optimal Transportation

In this subsection, we show how the results discussed in the previous sections can be restated in the more general context of *Optimal Transportation Theory.* We refer to [52] for a more complete discussion of this important research field and of the connection between the theory of *Monge-Ampère Equations* and *Optimal Transportation.*

Consider two measures η and ν defined over two measurable spaces X and Y respectively. The problem here is to find a measurable map $T : X \to Y$ such that $T_\sharp \eta = \nu$, namely such that

$$\nu(A) = \eta(T^{-1}(A)) \quad \forall\ A \subset Y,$$

and such that it minimizes a certain cost function c:

$$\int_X c(x,T(x))d\eta(x) = \min_{S_\sharp\eta=\nu}\int_X c(x,S(x))d\eta(x).$$

Here $c : X \times Y \to \mathbb{R}$ is the *cost function* and $S : X \to Y$ is a measurable map such that $S_\sharp\eta = \nu$.

Definition 2.2.18. *Suppose that $T : X \to Y$ and that $T_\sharp\eta = \nu$, then T is called* Transport Map*. Suppose moreover that T satisfies*

$$\int_X c(x,T(x))d\eta(x) = \min_{S_\sharp\eta=\nu}\int_X c(x,S(x))d\eta(x).$$

Then T is called Optimal Transport Map*.*

The problem of *Optimal Transportation* is in general very difficult, already in the simplest scenarios like the one of the *Euclidean Space* $X = Y = \mathbb{R}^n$ with the *Euclidean Distance* $c(x,y) = ||x-y||^2 := \Sigma_{i=1}^n|x_i - y_i|^2$ as *cost function*. We note also that not all the possible η and ν are admissible since not for all η and ν, it is possible to find at least one transport map, like for example in the case that one of the two measures η and ν is the *Dirac delta*, while the other is not.

In the cases that we have treated in the previous sections, we have that $X, Y \subset \mathbb{R}^n$ and that the measures are $\eta(dx) = f(x)dx$ and $\nu(dy) = g(y)dy$. Now, if $S : X \to Y$ is a sufficiently smooth transport map, we can rewrite the transport condition $S_\sharp\eta = \nu$ as a *Monge-Ampère Equation*. Suppose that $\phi : \mathbb{R}^n \to \mathbb{R}$ is a continuous function. Then, by duality, the condition $S_\sharp\eta = \nu$ can be rewritten as

$$\int_{\mathbb{R}^n}\phi(S(x))f(x)dx = \int_{\mathbb{R}^n}\phi(y)g(y)dy.$$

Now, as before, if we assume that $det(\nabla S(x)) \neq 0$ we can use Theorem 2.2.1 (the *Change of Variables Formula*) to get:

$$\int_{\mathbb{R}^n}\phi(S(x))f(x)dx = \int_{\mathbb{R}^n}\phi(y)g(y)dy = \int_{\mathbb{R}^n}\phi(S(x))g(S(x))|det(\nabla S(x))|dx.$$

This must be true for every *test function* $\phi(y)$ and therefore we have

$$f(x) = g(S(x))|det(\nabla S(x))|,$$

which leads again to the *Monge-Ampère Equation* as before. In this context, we can restate Theorem 2.2.2 as follows:

Theorem 2.2.19. *Suppose that X is a random variable and that $\eta(dx) = f_X(x)dx$, with $f_X(x)$ the pdf of X and*

$$\nu(dx) = f_X^{\frac{1}{\alpha}}(x)dx$$

its $\frac{1}{\alpha}$-counterpart. Then we can construct explicit Transport Maps *in the following cases:*

- *if η is the pdf of a* Multivariate Normal Distribution $X \sim MVN(0\Sigma)$ *and ν is its $\frac{1}{\alpha}$-counterpart, then $S(y) = (\sqrt{\alpha}y_1, \ldots, \sqrt{\alpha}y_n) = y\sqrt{\alpha}$.*
- *if η is the pdf of a* Multivariate Exponential Distribution $X \sim MVE(\lambda_1, \ldots, \lambda_n)$ *and ν is its $\frac{1}{\alpha}$-counterpart, then $S(y) = \alpha(y_1, \ldots, y_n) = \alpha y$.*
- *if η is the pdf of a* Cauchy Distribution $X \sim Cauchy(0)$ *and $\alpha = \frac{1}{2}$ and ν is its $\frac{1}{\alpha}$-counterpart, then $x = S^{-1}(y) = tan\left(\frac{y}{1+y^2} + arctan(y)\right)$.*

Similarly as in the context of the Monge Ampère Equation, the uniqueness of a *Transport Map* is not granted. To obtain it, one often tries to construct an *Optimal Transport Map* relatively to a certain *cost function* c with suitable hypotheses. For completeness, we recall here a result of Brenier [20] (see also [52]).

Theorem 2.2.20. *Let η and ν be two compactly supported probability measures on $\mathbb{R}^n$. If η is absolutely continuous with respect to the Lebesgue measure, then:*

- *There exists a unique solution T to the optimal transport problem with* cost function $c(x, y) = \frac{|x-y|^2}{2}$.
- *There exists a convex function $u : \mathbb{R}^n \to \mathbb{R}$ such that the optimal map T is given by $T(x) = \nabla u(x)$ for η-a.e. x.*

Furthermore if $\eta(dx) = f(x)dx$ and $\nu(dy) = g(y)dy$, then T is differentiable η-a.e. and

$$|det(\nabla T(x))| = \frac{f(x)}{g(T(x))}$$

for η-a.e. $x \in \mathbb{R}^n$.

Remark 2.2.21. *In the cases in which one can compute explicitly a* Transport Map T, *it would be interesting to find a criterion of uniqueness, and hence a* cost function c, *such that the explicit* Transport Map T *is actually an* Optimal Transport Map *for that c. A* *similar criterion could be then applied to* Transformation Theory *in order to choose a* natural *map that, at the level of the random variables, sends the random variable to its* $\frac{1}{\alpha}$*-counterpart.*

Remark 2.2.22. *We notice that, in both the cases of the* Multivariate Normal Distribution *and the* Multivariate Exponential Distribution*, we constructed the* Transport Maps $T(u) = \alpha^{1/2}u$ *and* $T(u) = \alpha u$ *respectively, for which there exist convex functions* K *such that* $T = \nabla K$*. For the* Multivariate Normal Distribution *we have* $K = \frac{1}{2}\alpha^{1/2}||u||^2$*, while for the* Multivariate Exponential Distribution *we have* $K = \frac{1}{2}\alpha||u||^2$ *with* $||u||^2 = \Sigma_{i=1}^{n}|u_i|^2$.

2.2.5 Some Numerics

In this subsection, we plan to illustrate graphically the transformations that we have constructed in the previous sections. As explained before, the transformation t can be interpreted as a *Transport Map.*

In the graph below, we can see that, while decreasing the values of α, the mass moves towards the mean of the distribution and this is in accordance with the *Law of Large Numbers.* We chose the length of the domain so that 0.995 of the mass of the distribution is included in the graph. Given a percentage p, the *quantiles* $Q(p)$ cannot be computed explicitly for the *Normal Distribution* (there are of course tables), but there is an explicit formula for both the *Exponential Distribution* $Q_e(p) = -log(1-p)$ and for the *Cauchy distribution* $Q_c(p) = tan\left(\pi(p-0.5)\right)$. Here p stays for the percentage or *plotting point.*

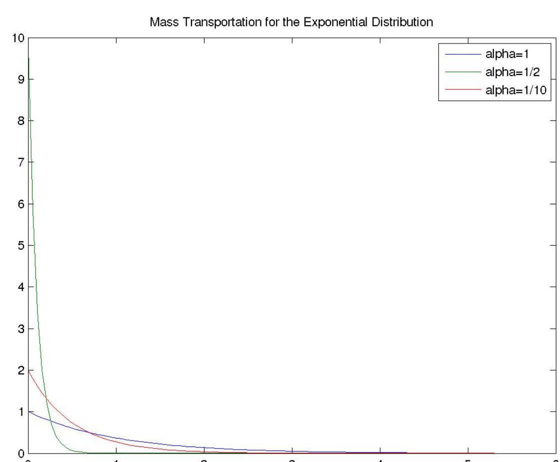


Figure 2.1: Mass Transportation in The Exponential Distribution

```
%Exponential Distributions: ya=1/a*exp(x/a)
x=linspace(0,5.3,100); %Area inside 99.5
y1=exp(-x);
y2=10*exp(-10*x);
y10=2*exp(-2*x);
plot(x,y1,x,y2,x,y10)
title('Mass Transportation for the Exponential Distribution')
legend('alpha=1','alpha=1/2', 'alpha=1/10')
```

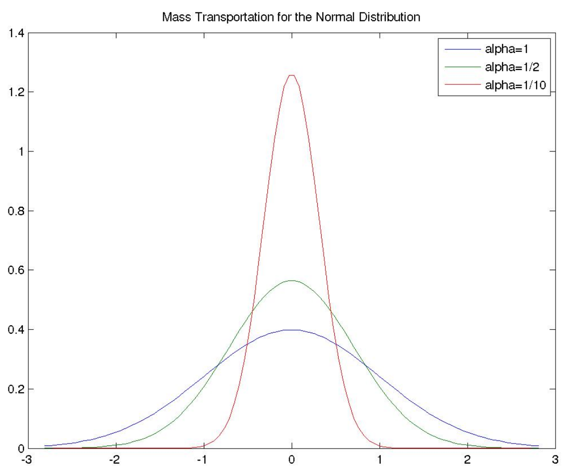


Figure 2.2: Mass Transportation in The Normal Distribution

```
%Normal Distributions: ya=1/(2*pi*a)^(1/2)*exp(-x^2/(2*a))
x=linspace(-2.81,2.81,100); %Area inside 99.5
y1=1/(2*pi*1).^(1/2)*exp(-x.^2/(2*1));
y2=1/(2*pi*(1/2)).^(1/2)*exp(-x.^2/(2*(1/2)));
y10=1/(2*pi*(1/10)).^(1/2)*exp(-x.^2/(2*(1/10)));
plot(x,y1,x,y2,x,y10)
title('Mass Transportation for the Normal Distribution')
legend('alpha=1','alpha=1/2', 'alpha=1/10')
```

```
%Cauchy Distributions:
x=linspace(-127.32,127.32,100); %Area inside 99.5
y1=1/pi*(1+x.^2).^(-1);
```

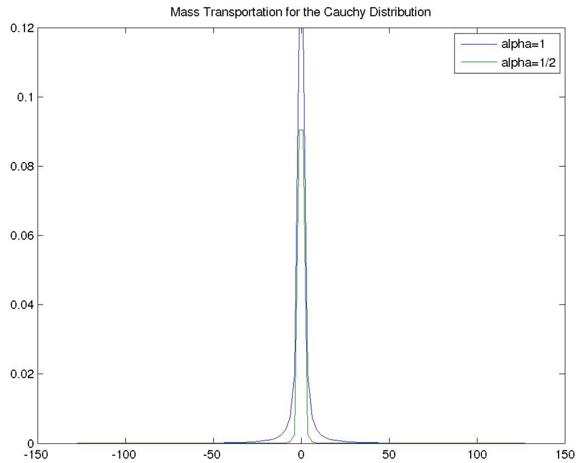


Figure 2.3: Mass Transportation in The Cauchy Distribution

```
y2=2/pi*(1+x.^2).^(-2);
plot(x,y1,x,y2)
title('Mass Transportation for the Cauchy Distribution')
legend('alpha=1','alpha=1/2')
```

We notice that indeed, when the parameter $\alpha \to 0$, the distributions tend weakly to the *Dirac Delta*. For the sake of illustration, we verify this property in the case of the *Exponential Distribution*. Suppose that $X \sim exp(\lambda)$, then $f_X(x;\lambda) = \lambda e^{-x\lambda}$ and so $f^{\frac{1}{\alpha}}(x;\lambda) = \frac{\lambda^{\frac{1}{\alpha}} e^{-x\frac{\lambda}{\alpha}}}{\lambda^{1/\alpha-1}\alpha}$. Suppose also that $\phi(x)$ is a continuous function on $[0,+\infty]$ with compact support inside $[0,+\infty]$. Then, by duality,

$$\int_0^{+\infty} f^{\frac{1}{\alpha}}(x;\lambda)\phi(x)dx = \int_0^{+\infty} \frac{\lambda}{\alpha} e^{-x\frac{\lambda}{\alpha}}\phi(x)dx = \int_0^{+\infty} e^{-x}\phi(\alpha\lambda^{-1}x) \to \phi(0),$$

as $\alpha \to 0$ and λ fixed (this actually works also when $\lambda = \lambda(\alpha)$ and α goes to zero faster than λ). And so this means that

$$f^{\frac{1}{\alpha}}(x;\lambda) \xrightarrow{w} \delta_{x=0},$$

as $\alpha \to 0$. This phenomenon becomes even more evident through an animation, that can be performed by the following simple code in MATLAB. Here, we concentrate in particular on the cases of the *Exponential Distribution* and *Normal Distribution*, but similar animations can be performed for other distributions.

```
%ANIMATION EXPONENTIAL
FramesNumbers= 100;
frames = moviein(FramesNumbers);
x = 0 : .01 : 10;
p = 1;
for i = 1 : FramesNumbers
    p = p +1;
    y = p.*exp(-x.*p);
    plot(x, y);
    title('Mass Transportation in The Exponential Distribution')
    frames(:, i) = getframe;
end
save frames
```

```
%ANIMATION NORMAL
FramesNumbers= 100;
frames = moviein(FramesNumbers);
x = -10 : .01 : 10;
p = 1;
for i = 1 : FramesNumbers
    p = p +1;
    y = 1/(2*pi*p^(-1))^(1/2).*exp(-x.^2*p/2);
    plot(x, y);
    title('Mass Transportation in The Normal Distribution')
    frames(:, i) = getframe;
end
save frames
```

Remark 2.2.23. *The code for the animation has been built, taking inspiration from [3].*

2.3 Applications to the Consensus Monte Carlo Algorithm

We conclude this chapter with some final comments, where we mention that what we called $\frac{1}{\alpha}$-*Characteristic Functions* appear implicitly in the context of the *Consensus Monte Carlo Algorithm* (see [105]). We briefly explain where and how and leave more detailed comments to later on, when in Section 3.3 we discuss the *Simpson's Paradox* in this context.

Big data can be seen as data that, for one reason or another, is too big to efficiently process on a single machine. Some of the main problems are processor, memory, or disk bottlenecks and often they can be eliminated just by distributing the data across different machines. This is computationally expensive and so there is need of efficient algorithms to reduce the costs.

The *Consensus Monte Carlo* is a possible solution and works in the following way. First, it breaks the data into *shards*, then it sends each *shard* to a different machine which independently runs a *Monte Carlo Algorithm* from a posterior distribution given its own data, and then combine the posteriors (*consensus*).

Let $\mathbf{y}$ represent the full data set, and $\mathbf{y}_s$ be the fraction of the data set corresponding to *shard* s. Moreover, let θ be the model parameters. Then, the posterior distribution of the system can be written as

$$p(\theta, y) \propto \prod_{s=1}^{S} p(y_s|\theta)p(\theta)^{1/S},$$

where S represents the total number of *shards*. The prior distribution $p(\theta)$ is broken into S components $p(\theta)^{1/S}$ to preserve the total amount of prior information in the system and totally resembles our $1/\alpha$-*Probabilities*, with $\alpha = S$.

Our results can be then seen as a theoretical analysis of some of the key objects appearing in the algorithm in [105], also in the case of non-Gaussian posteriors, which require some extra theoretical work according to the authors of [105].

2.4 Concluding Remarks

This chapter discussed the characterization of distributions using properties related to optimal mass transportation. The methods used, which have strong connection with Fourier analysis can be applied to do statistical inference for distributions that do not possess good regularity, decay or integrability properties, like the Cauchy distribution. As a possible application, we have discussed some topics in Big Data analysis and in particular the Consensus Monte Carlo Algorithm. Further characterizations of probability distributions might be available using these techniques and might be good to work on them in the future.

Chapter 3

The Simpson's Paradox

In this chapter, we study the Simpson's Paradox. The *Simpson's Paradox* is the phenomenon that appears in some datasets, where subgroups with a common trend (say, all negative trend) show the reverse trend when they are aggregated (say, positive trend).

In the first section, we give a brief introduction of the problem and the second section, we prove that the *Simpson's Paradox* occurs also in an unconventional settings, like the one of Quantum Mechanics.

3.1 The Ubiquity of the Simpson's Paradox

As just said, the *Simpson's Paradox* is the phenomenon that appears in some datasets, where subgroups with a common trend (say, all negative trend) show the reverse trend when they are aggregated (say, positive trend). Even if this issue has an elementary mathematical explanation, it has a deep statistical significance.

In this section, we discuss basic examples in arithmetic, geometry, linear algebra, statistics, game theory, gender bias in university admission and election polls, where we describe the appearance or absence of the *Simpson's Paradox*.

3.1.1 Introduction and Motivation

In probability and statistics, the *Simpson's paradox* (called also *Yule-Simpson effect*) is a paradox in which a trend that appears in different groups of data disappears when these groups are combined, while the reverse trend appears for the aggregate data. This effect is often encountered in social sciences, psychology, medical sciences in general, ecology, theoretical statistics and several other fields [121].

The problem of the occurrence of this paradox is a very old one and dates back to the 19th century. The first author which treated this topic has been Pearson [95], who pointed out the occurrence of the paradox, while studying correlation measures for continuous, non-categorical data. Subsequently, Yule [125] [126] pointed out that "a pair of attributes does not necessarily exhibit independence within the universe at large even if it exhibits independence in every subuniverse". See also [127].

Simpson first described this phenomenon in a paper [109], where he considered a 2x2x2 contingency table with attributes A, B, and C and illustrated the paradox using a heuristic example of clinic patients. In the example, patients received treatment or no treatment. When the data were examined by gender (subpopulations), one was led to conclude that both males and females responded favorably to the treatment, compared to those who did not receive the treatment. However, when the data were aggregated (population), there seemed to not be anymore any association between the use of the treatment and the survival time. See [57] for more details.

The name *Simpson's paradox* was given to this phenomenon by Blyth [17]. Since Simpson did not actually discover this statistical paradox, some authors prefer to call it *reversal paradox* or *amalgamation paradox*. Good and Mittal [59] studied deeply how to avoid the paradox and gave necessary and sufficient conditions on the contingency tables and on the sampling experiments in order to avoid the paradox. See Subsection 3.2.2.3 for more details.

One popular example of the occurrence of the *Simpson's paradox* is the Berkeley Gender Bias Admission Problem [15]. Here is the story.

In 1973, the Associate Dean of the graduate school of the University of California Berkeley worried that the university might be sued for sex bias in the admission process [15]. In fact, looking at the admission rates broken down by gender (male or female), we have the following contingency table:

Applicants	Admitted	Deny
Female	1494	2827
Male	3738	4704

The Chi-square statistics for this test has one degrees of freedom with value $\chi^2 = 111.25$ and corresponding p-value basically $= 0$, while the Chi-square statistics with Yates continuity correction for this test has a value of $\chi^2 = 110.849$ and corresponding p-value again approximately 0 (precision order 10^{-26}). A naïve conclusion would be that men were much more successful in admissions than women, which would clear be understood as a bad episode of gender bias. At that point, Prof. P.J.Bickel from the Department of Statistics of Berkeley, was asked to analyse the data.

In a famous paper [15] with E.A.Hammel and J.W.O'Connell, P.J.Bickel studied the problem in detail. Graduate departments have independent admissions procedures and so they are autonomous for taking decisions in the graduate admission process. A further division in subgroups does not find a real counterpart in the structure of Berkeley's system. The analysis of the data, performed department by department, produces the following table:

Dpt	Male Applications	Male Admissions	Female Applications	Female Admissions
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	191	28%	393	24%

As Bickel, Hammel and O'Connell say in [15], "The proportion of women applicants tends to be high in departments that are hard to get into and low in those that are easy to get into" and it is even more evident in departments with a large number of applicants. The examination of the aggregate data was showing a misleading pattern of bias against female applicants. However, if the data are properly pooled, and taking into consideration the tendency of women to apply to departments that are more competitive for either genders, there is a small but statistically significant bias in favour of women. The authors concluded that "Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation" [15]. This episode is one of the most celebrated real examples of what is called *Simpson's Paradox*: the trend of aggregated data might be reversed in the pooled data.

Note that the *Simpson's Paradox* is not confined to the discrete case, but it can appear also in the continuous case. Even if less famous, we want to mention the following example which has been discussed on the New York Times recently [91]. Still today, the *Simpson's*

Paradox can be a source of confusion and misinterpretation of the data.

An article of the journalist F.Norris [91] raised the concerns of readers, because of the following apparently paradoxical result. F.Norris analysed the variation of the US wage over time. Accordingly to the statistics, from 2000 to 2013, the median US wage (adjusted for inflation) has risen of about 1%, if the median is computed on the full sample. However, if the same sample is broken down into four educational subgroups, the median wage (adjusted for inflation) of each subgroup decreased. The percentages of variation for each subgroup are summarized in the following table:

Group	Median Change
Total	+0.9%
High School Dropouts	-7.9%
High School Graduates, No college	-4.7%
Some College	-7.6%
Bachelor's or Higher	-1.2%

Here, the reason of the reversal is that the relative sizes of the groups changed greatly over the period considered. In particular, there were more well-educated and so higher wage people in 2013 than in 2000.

In both the cases described above (discrete and continuous, respectively), the variables involved in the paradox are confounded by the presence of another variable (department and level of education, respectively).

Remark 3.1.1. *For more examples we refer to [78], [121] and the references therein.*

The problem of the occurrence of this paradox was considered already in the 19th century. The first author which treated this topic has been Pearson [95], followed by the contributions of Yule [125] [127] and Simpson [109]. In this section, we outline that the *Simpson's Paradox* is not confined to statistical problems, but it is ubiquitous in science. We give a series of formal definitions in Subsection 3.1.2. In Subsection 3.1, we show the ubiquity of the *Simpson's Paradox* in several areas of technical and social sciences and we also give some examples of its occurrence.

3.1.2 Measures of Amalgamation

In this section, we give the definition and some popular examples of *Measures of Amalgamation*. For more details, we refer to [59].

3.1.2.1 Definitions

First, we define the *Process of Amalgamation* of contingency tables $\mathbf{t}_i, i = 1, \ldots, n$.

Definition 3.1.2. *Let* $\mathbf{t}_i = [a_i, b_i; c_i, d_i]$, $i = 1, \ldots, l$ *be* 2×2 **contingency tables** *corresponding to the* i*-th of* l *mutually exclusive sub-populations, with* $a_i b_i c_i d_i \neq 0$*. Let* $N_i = a_i + b_i + c_i + d_i$ *denote the sample size for the* i*-th sub-population and let* $N = N_1 + \cdots + N_l$ *be the total sample size of the population. If the* n *tables are added together, the process is called* **Amalgamation***. We obtain a table* $\mathbf{T} := [A, B; C, D] := [\Sigma_{i=1}^l a_i, \Sigma_{i=1}^l b_i, \Sigma_{i=1}^l c_i, \Sigma_{i=1}^l d_i]$*, where* $A + B + C + D = N$*.*

After having amalgamated a group of contingency tables, we can define the *Measure of Amalgamation*.

Definition 3.1.3. *Let* $M_{p\times p}$ *be the set of all* $p \times p$ *contingency tables. A function* $\alpha : M_{p\times p} \to \mathbb{R}$ *is called* ***Measure of Amalgamation****.*

Given the definition of *Measure of Amalgamation*, we can formally define the *Simpson's Paradox*.

Definition 3.1.4. *We say that the* **Simpson's Paradox** *occurs for the* **Measure of Amalgamation** α *if*

$$\max_i \alpha(\mathbf{t}_i) < \alpha(\mathbf{T}) \text{ or } \min_i \alpha(\mathbf{t}_i) > \alpha(\mathbf{T}),$$

with α *defined as in Definition 3.1.3 and* $\mathbf{T} = \mathbf{t}_1 + \mathbf{t}_2$*.*

We fix some terminology that we are going to use in the list of examples below in the context of contingency tables (see [59]). Sampling Procedure I, called also *Tetranomial Sampling*, is performed when we sample at random from a population. Sampling Procedure II_R (respectively II_C), called also *Product-Binomial Sampling*, is performed when the row totals (respectively columns) is fixed and we sample until this marginal totals are reached. Sampling Procedure III controls both row and column totals.

3.1.2.2 Examples

Consider the 2×2 contingency table $\mathbf{t} = [a, b; c, d]$, given by

$\mathbf{t} =$

	S	not S
T	a	b
not T	c	d

The following are popular examples of *Measures of Amalgamation* (see [59]).

- The *Pierce's measure*:
$$\pi_{Pearce}(\mathbf{t}) = \frac{a}{a+b} - \frac{c}{c+d}.$$
Under *Tetranomial Sampling* and *Product-Binomial Sampling* with row fixed, this measure becomes
$$\pi_{Pearce} = \hat{P}(S|T) - \hat{P}(S|\bar{T}).$$
It compares the estimated probability of an effect S under treatment and the estimated probability of an effect S without any treatment (row categories are considered to be the "causes" of the column categories).

- The *Yule's measure* is given by the formula:
$$\pi_{Yule}(\mathbf{t}) = \frac{ad-bc}{N^2}.$$
It compares the frequency a/N with respect to its estimated expected frequency under independence of rows and columns. In fact:
$$\pi_{Yule}(\mathbf{t}) = \frac{ad-bc}{N^2} = \frac{a}{N} - \frac{(a+b)(a+c)}{N^2} = \hat{P}(S\cap T) - \hat{P}(S)\hat{P}(T),$$
since $N = a+b+c+d$.

- The *Odds Ratio* is arguably the most popular measure of amalgamation:
$$\pi_{Odds}(\mathbf{t}) = \frac{ad}{bc}.$$
The *Odds Ratio* is the ratio between the estimated probability of success and the estimated probability of failure, given a treatment or a no-treatment. In fact
$$\pi_{Odds}(\mathbf{t}) = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{\frac{a/(a+b)}{b/(a+b)}}{\frac{c/(c+d)}{d/(c+d)}} = \frac{\hat{P}(S|T)/\hat{P}(\bar{S}|T)}{\hat{P}(S|\bar{T})/\hat{P}(\bar{S}|\bar{T})}.$$

- The *Weight of Evidence* is given by:
$$\pi_{Weight_C}(\mathbf{t}) = \log\frac{a(b+d)}{b(a+c)}.$$
Under *Tetranomial Sampling* or column fixed *Product-Binomial Sampling*,
the *Weight of Evidence* represents the logarithm of the estimated *Bayes factor* in

favour of S, knowing that the treatment was T, namely:

$$\pi_{Weight_C} = \log \frac{\hat{P}(T|S)}{\hat{P}(T|\bar{S})}.$$

- The *Causal Propensity*:

$$\pi_{Causal}(\mathbf{t}) = \log \frac{d(a+b)}{b(c+d)},$$

  under *Tetranomial Sampling* or *Product-Binomial Sampling* with row fixed, represents the estimated propensity of T causing S rather than $\bar{S}$:

$$\pi_{Causal}(\mathbf{t}) = \log \frac{\hat{P}(\bar{S}|\bar{T})}{\hat{P}(\bar{S}|T)}.$$

3.1.3 The *Simpson's Paradox* appears not just in statistics

In this section, we give very basic examples of the appearance of the *Simpson's Paradox* in fields different from statistics. In particular, we give examples in arithmetic, geometry, statistics, linear algebra, game theory and election polls.

- **Arithmetic**: There exist quadruplets $a_1, b_1, c_1, d_1 > 0$ and $a_2, b_2, c_2, d_2 > 0$ such that $a_1/b_1 > c_1/d_1$ and $a_2/b_2 > c_2/d_2$ but $(a_1 + a_2)/(b_1 + b_2) < (c_1 + c_2)/(d_1 + d_2)$. Example: $(a_1, b_1, c_1, d_1) = (2, 8, 1, 5)$ and $(a_2, b_2, c_2, d_2) = (4, 5, 6, 8)$. In this case, the *Measure of Amalgamation* is given by:

$$\pi(\mathbf{t}) = \frac{a}{b} - \frac{c}{d} = \frac{ad - bc}{bd}.$$

  If we consider the contingency tables

$$\mathbf{t}_1 = [a_1, b_1; c_1, d_1]$$

  and

$$\mathbf{t}_2 = [a_2, b_2; c_2, d_2]$$

  and the amalgamated one:

$$\mathbf{T} = [a_1 + a_2, b_1 + b_2; c_1 + c_2, d_1 + d_2],$$

  we have that:

$$\max_{i=1,2} \pi(\mathbf{a_i}) < 0 < \pi(\mathbf{T})$$

and so we have the *Simpson's Paradox*, accordingly to Definition 3.1.4.

- **Geometry**: Even if a vector v_1 has a smaller slope than another vector w_1, and a vector v_2 has a smaller slope than a vector w_2, the sum of the two vectors $v_1 + v_2$ can have a larger slope than the sum of the two vectors $w_1 + w_2$. Example: take $w_1 = (a_1, b_1)$, $v_1 = (c_1, d_1)$, $w_2 = (a_2, b_2)$, $v_2 = (c_2, d_2)$. The same *Measure of Amalgamation* of the previous example makes the game here as well.

- **Statistics**: A positive/negative trend of two separate subgroups might reverse when the subgroups are combined in one single group. This happens in both the discrete and continuous case. We gave examples of this in the introduction, with the Berkeley Gender Bias (discrete) case and the "time vs US wage" case (continuous).

- **Linear Algebra** There exists $A_1, A_2 \in Mat_{n\times n}$ such that

$$det(A_1) > 0, \quad det(A_2) > 0, \quad but \quad det(A_1 + A_2) < 0.$$

  Consider for example $A_1 = \mathbf{t}_1$ and $A_2 = \mathbf{t}_2$, as above.

- **Game Theory**: The *Prisoner's Dilemma* shows why two players A and B might decide to not cooperate, even if it appears that, for both of them, it is more convenient to cooperate. If both A and B cooperate, they both receive a reward p_1. If B does not cooperate while A cooperates, then B receives p_2, while A receives p_3. Similarly, if vice versa. If both A and B do not cooperate, their payoffs are going to be p_4. To get the *Simpson's Paradox*, the following must hold:

$$p_2 = a_2/b_2 > p_1 = c_2/d_2 > p_4 = a_1/b_1 > p_3 = c_1/d_1.$$

  Here $p_2 > p_1$ and $p_4 > p_3$ imply that it is better to not cooperate for both A and B both given the fact that the other player does or does not cooperate (*Nash Equilibrium*). Note that, if we use these quadruplets for the table of rewards, we get for the rewards of player A:

Rewards for A	B cooperates	B does not
A cooperates	p_1	p_3
A does not	p_2	p_4

  and for the rewards of player B:

Rewards for B	B cooperates	B does not
A cooperates	p_1	p_2
A does not	p_3	p_4

Using the values in our examples, we get for the rewards of player A:

Rewards for A	B cooperates	B does not
A cooperates	0.75	0.2
A does not	0.8	0.25

and for the rewards of player B:

Rewards for B	B cooperates	B does not
A cooperates	0.75	0.8
A does not	0.2	0.25

Note that this implies that both players A and B are pushed, for personal convenience, to not cooperate, independently of what the other player does, but end up getting a worse reward than if they would have both cooperated. In fact, the *amalgamated* contingency table, gives:

Rewards for A+B	B cooperates	B does not
A cooperates	1.5	1
A does not	1	0.5

that prizes the decision of cooperation. The *Measure of Amalgamation* considered here can be thought in the form of an *Utility Function*, such as:

$$U_A(a,b) = p_1ab + p_3a(1-b) + p_2b(1-a) + p_4(1-a)(1-b)$$

and

$$U_B(a,b) = p_1ab + p_2a(1-b) + p_3b(1-a) + p_4(1-a)(1-b).$$

Here $a = 1$, means that A cooperates, while $a = 0$ means that A does not. Similarly for B. Note that, under the conditions on p_1, p_2, p_3 and p_4 mentioned above, the Utility is bigger for the choice of not cooperation for both A and B, given any decision taken by the other player. In fact,

$$p_1 = U_A(1,1) < U_A(0,1) = p_2$$

and

$$p_3 = U_A(1,0) < U_A(0,0) = p_4$$

and analogously for U_B. However, when we combine the utilities, we get *Utility Function*

$$U_{A+B}(a,b) = 2p_1ab + (p_2 + p_3)a(1-b) + (p_3 + p_2)b(1-a) + 2p_4(1-a)(1-b).$$

This utility is always bigger for cooperation, if we require $2p_4 < p_2 + p_3 < 2p_1$, as we chose in our example. In fact:

$$2p_4 = U_{A+B}(0,0) < U_{A+B}(1,0) = p_2 + p_3 = U_{A+B}(0,1) < 2p_1 = U_{A+B}(1,1).$$

In this way, we have restated the *Prisoner's Dilemma* in the context of the *Simpson's Paradox*.

- **Election Polls**: Suppose candidates T and C run for elections in two states $State_1$ and $State_2$. Suppose that candidate T and C receive in $State_1$ a percentage of votes:

$$\%votes\ for\ T = \frac{a}{b} > 1 - \frac{a}{b} = \%votes\ for\ C$$

and that candidate T and C receive in $State_2$ a percentage of votes:

$$\%votes\ for\ T = \frac{c}{d} > 1 - \frac{c}{d} = \%votes\ for\ C.$$

Is it possible that overall candidate C receives a higher percentage of votes? Clearly, this is not possible because $\frac{a}{b} > 1 - \frac{a}{b}$ implies $a > 0.5b$ and $\frac{c}{d} > 1 - \frac{c}{d}$ implies $c > 0.5d$ and so

$$0.5b + 0.5d < a + c,$$

which implies

$$\frac{a+c}{b+d} > 0.5$$

and so

$$\frac{a+c}{b+d} > 1 - \frac{a+c}{b+d}.$$

In this case, we do not have any paradox and this is related to the fact that there is an extra constraint on the construction of the contingency table. Note that since the set of real numbers for which these inequalities hold is an open set, the inclusion of a not strong third candidate will not change the situation. What happens if the third candidate is as strong as T and C? We plan to work on this and the problem of gerrymandering in the future.

3.2 The Simpson's Paradox in Quantum Mechanics

In this section, we give some results about the occurrence of the *Simpson's Paradox* in Quantum Mechanics.

In particular, we prove that the *Simpson's Paradox* occurs for solutions of the *Quantum Harmonic Oscillator* both in the stationary case and in the non-stationary case. In the non-stationary case, the *Simpson's Paradox* is persistent: if it occurs at any time $t = \tilde{t}$, then it occurs at any time $t \neq \tilde{t}$. Moreover, we prove that the *Simpson's Paradox* is not an isolated phenomenon, namely that, close to initial data for which it occurs, there are lots of initial data (a open neighborhood), for which it still occurs. Differently from the case of the *Quantum Harmonic Oscillator*, we also prove that the paradox appears (asymptotically) in the context of the *Nonlinear Schrödinger Equation*, but at intermittent times.

We conclude with a discussion about the likelihood of incurring in the *Simpson's Paradox* in Quantum Mechanics, some numerical examples and some final considerations and open problems.

3.2.1 Introduction and Motivation

In this section, we show that the occurrence of the *Simpson's Paradox* can be detected in a completely different setting with respect to the ones more commonly studied. In particular, we will show that the phenomenon appears in *quantum mechanics* too and more precisely in the context of the *Quantum Harmonic Oscillator* and the *Nonlinear Schrödinger Equation.*

Very few papers in the literature treat the *Simpson's Paradox* related to problems in *Quantum Mechanics*. At our knowledge, the only ones available are the fast track communication by Paris [93], an experimental result by Cialdi and Paris [31], the preprint by Shi [108] and our work, which is the first that connects the *Simpson's Paradox* to Partial Differential Equations and Infinite Dimensional Dynamical Systems.

We consider the *Linear Schrödinger Equation* in the presence of a *Harmonic Potential*:

$$i\hbar\frac{\partial}{\partial t}\psi(t,x) = -\frac{\hbar^2}{2m}\Delta\psi(t,x) + \frac{1}{2}m\omega^2|x|^2\psi(t,x) \quad (t,x) \in (0,+\infty)\times\mathbb{R}^n. \tag{3.1}$$

Here $i = \sqrt{-1}$ is the *complex unit*, $\hbar$ is the *Planck Constant*, m represents the *mass* of a particle and ω is the *angular velocity.* For this equation, we prove that, with a proper choice of the initial datum and the proper choice of the *measure of amalgamation*, the *Simpson's Paradox* can occur.

Theorem 3.2.1. *[Existence of the* Simpson's Paradox*]*
Consider equation (3.1) for every spatial dimension $n \geq 1$. Then, for every $m > 0$, $\omega > 0$, there exists a set of parameters $(x_i(t), \gamma_i(t), v_i(t))$ for $i = 1, \dots, 4$, such that the following is true. If we consider an initial datum of the form $\psi(0,x) = \Sigma_{i=1}^4 \psi_i(0,x)$ with $\psi_i(0,x)$ the soliton solutions as in Theorem 3.2.8 below, namely such that

$$\psi_i(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_i(t)+\gamma_i(t)+\frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x-x_i(t)|^2},$$

then the Simpson's Paradox *occurs in the following cases:*

- *in the* stationary case*, namely when $v_i(t) = 0$ and $x_i(t) = x_i$ for every t both when $\gamma_i = \gamma_j$ for every $1 \leq i, j \leq 4$ and when $\gamma_i \neq \gamma_j$ $1 \leq i, j \leq 4$, $i \neq j$.*
- *in the* non-stationary case*, if there exists $t_0 \in \mathbb{R}$ such that the* Simpson's Paradox *occurs at t_0, then the* Simpson's Paradox *occurs at any t_1 with $t_1 \neq t_0$ too.*

For the choice of the *measure of amalgamation* for which the paradox occurs, we refer to Subsection 3.2.2.3 and the following.

Remark 3.2.2. *As it will become evident in the proof of the theorems, this phenomenon appears at both* semiclassical *and* anti-semiclassical *scales. See Section 3.2.3 and Remark 3.2.19.*

Remark 3.2.3. *The fact that in the* non-stationary case *the* Simpson's Paradox *when it appears at a time t_0, it persists for any time $t_1 \neq t_0$ is due to the fact that the* measure of amalgamation *that we choose is constant in time for the type of solutions that we are considering. See Section 3.2.3 below.*

Remark 3.2.4. *This theorem can be seen as a criterion to avoid the paradox in the spirit of [59]. In fact it is enough to avoid the paradox at time $t = \tilde{t}$ for some $\tilde{t} \in \mathbb{R}$ to avoid the paradox forever. This is a good property, because, the possibility of building contingency tables which avoid the paradox, simplifies the interpretation of the data.*

This phenomenon is not an isolated phenomenon, namely close to solutions which exhibit the *Simpson's Paradox*, it is full of solutions which do the same. This is in fact the content of the following theorem.

Theorem 3.2.5. *[Stability of the* Simpson's Paradox*]*
Suppose that there exists a set of parameters $(x_i(t), \gamma_i(t), v_i(t))$ *for* $i = 1, \ldots, 4$ *such that the* Simpson's Paradox *occurs in the stationary case. Then, there exists* $r > 0$ *such that, for every* $(\tilde{x}_i(t), \tilde{\gamma}_i(t), \tilde{v}_i(t))$ *for* $i = 1, \ldots, 4$ *inside* $B_r((x_i(t), \gamma_i(t), v_i(t)), \quad i = 1, \ldots, 4)$, *the* Simpson's Paradox *still occurs for initial data as in Theorem 3.2.1. Moreover, if the* Simpson's Paradox *occurs for a* $\psi(t,x)$ *at a certain time* $\tilde{t}$*, then there exists and open ball in* $\Sigma := L^2(\mathbb{R}^n, dx) \cap L^2(\mathbb{R}^n, |x|^2 dx)$ *such that the* Simpson's Paradox *still occurs for any* $\bar{\psi}(\tilde{t}, x) = \psi(\tilde{t}, x) + w(\tilde{t}, x)$ *with* $w(\tilde{t}, x) \in \Sigma$ *and the same time* $\tilde{t}$*.*

Moreover, the *Simpson's Paradox* can be detected also in the case of a *Focusing Nonlinear Schrödinger Equation* (NLS) asymptotically in time, at least in the case of spatial dimension $n = 1$. In this case though, the paradox is *intermittent*. See Section 3.2.8 for more details.

Theorem 3.2.6. *[Nonlinear case] Consider equation*

$$\begin{cases} i\frac{\partial}{\partial t}\psi(t,x) = -\Delta_x\psi(t,x) - |\psi(t,x)|^{p-1}\psi(t,x), & (t,x) \in \mathbb{R} \times \mathbb{R}^n \\ \psi(0,x) = \psi_0(x), \end{cases}$$

where $n = 1$*,* $1 < p < 1 + \frac{4}{n}$ *is the* L^2-subcritical exponent*. Then, there exist initial data* $\psi_0(x)$ *in the form of* Multi-Solitons *(see 3.2.10 below and [88]) for which there exists* $\tilde{t} \gg 1$ *such that the* Simpson's Paradox *occurs at time* $\tilde{t}$*.*

Remark 3.2.7. *For some comments about the nonlinear case in dimension* $n > 1$*, we refer to Section 3.2.8.*

The remaining part of the section is organized as follows. In Subsection 3.2.2, we collect some preliminary results, like the construction of the general moving soliton for the *Harmonic Oscillator*, the properties of the soliton and the construction of the multi-soliton for the *Nonlinear Schrödinger Equation* and the discussion about the choice of the *measure of amalgamation*. In Subsection 3.2.3, we first give the proof of Theorem 3.2.1 in the case of dimension $n = 1$, both stationary and non-stationary case, and then we adapt it to the case of spatial dimension $n \geq 2$. In Subsection 3.2.4, we give the complete proof of Theorem 3.2.5, while, in Subsection 3.2.5, we give the proof of Theorem 3.2.6. Finally, we give some final considerations and talk about open problems in Subsection 3.2.8.

3.2.2 Preliminaries

In this subsection, we collect some preliminary results that we will need in the proofs of the theorems. We first construct *The Moving Soliton Solution* of the *Quantum Harmonic Oscillator*, we discuss the properties of the soliton and the construction of the multi-soliton for the *Nonlinear Schrödinger Equation*. In the last subsubsection, we discuss the chosen *Measure of Amalgamation*.

3.2.2.1 The Moving Soliton Solution of the Quantum Harmonic Oscillator

We look for solutions which represent *moving solitons*, namely solutions of the form:

$$\psi(t,x) = u(\mathbf{x}-\mathbf{x}(t))e^{i[\mathbf{x}\cdot\mathbf{v}(t)+\gamma(t)+\frac{\omega t}{2}]} \tag{3.2}$$

for a certain choice of $u(\mathbf{x})$, $\mathbf{x}(t)$, $\mathbf{v}(t)$ and $\gamma(t)$.

Theorem 3.2.8. *There exists a solution of equation (3.1) of the form (3.2) with the following conditions on* $u(\mathbf{x})$, $\mathbf{x}(t)$, $\mathbf{v}(t)$ *and* $\gamma(t)$*:*

- *the* profile $u(\mathbf{x})$ *satisfies the equation*

$$-\frac{\hbar^2}{2m}\Delta u(\mathbf{x}) + \frac{1}{2}m\omega^2|\mathbf{x}|^2 u(\mathbf{x}) + \frac{\omega}{2}u(\mathbf{x}) = 0, \quad \mathbf{x}\in\mathbb{R}^n; \tag{3.3}$$

- *the* position vector $\mathbf{x}(t)$ *and the* velocity vector $\mathbf{v}(t)$ *satisfy the following system of ODE*

$$\begin{cases} \dot{\mathbf{x}}(t) = \frac{\hbar}{m}\mathbf{v}(t) \\ \dot{\mathbf{v}}(t) = -\frac{m}{\hbar}\omega^2\mathbf{x}(t) \end{cases} \tag{3.4}$$

- *the* complex phase $\gamma(t)$ *is such that*

$$\dot{\gamma}(t) = \frac{1}{\hbar}\mathcal{L}(\mathbf{x}(t),\dot{\mathbf{x}}(t);m,\omega),$$

where $\mathcal{L}(\mathbf{x}(t),\dot{\mathbf{x}}(t);m,\omega) := \frac{1}{2}m|\dot{\mathbf{x}}(t)|^2 - \frac{1}{2}m\omega^2|\mathbf{x}(t)|^2$ *is the* Lagrangian *of the system of ODEs (3.4).*

Proof. The strategy here is to take a general solution of the form

$$\psi(t,x) = u(\mathbf{x}-\mathbf{x}(t))e^{i[\mathbf{x}\cdot\mathbf{v}(t)+\gamma(t)+\frac{\omega t}{2}]},$$

plug this *ansatz* into

$$i\hbar\frac{\partial}{\partial t}\psi(t,\mathbf{x}) = -\frac{\hbar^2}{2m}\Delta\psi(t,\mathbf{x}) + \frac{1}{2}m\omega^2|\mathbf{x}|^2\psi(t,\mathbf{x}) \quad (t,\mathbf{x}) \in (0,+\infty)\times\mathbb{R}^n$$

and then choose the *profile* $u(\mathbf{x})$ and the *parameters* $\mathbf{x}(t)$, $\mathbf{v}(t)$ and $\gamma(t)$ such that the equation (3.1) is satisfied. In the following, we will use the variable $\mathbf{y}$ defined by $\mathbf{y} := \mathbf{x} - \mathbf{x}(t)$. We start by computing all the derivatives involved in (3.1). First, the time derivative:

$$\frac{\partial}{\partial t}\psi(t,\mathbf{x}) = \left[-\dot{\mathbf{x}}\cdot\nabla_y u(\mathbf{y}) + iu(\mathbf{y})\left(\dot{\mathbf{v}}(t)\cdot\mathbf{y} + \dot{\mathbf{v}}(t)\cdot\mathbf{x}(t) + \dot{\gamma}(t) + \frac{m\omega}{\hbar}\right)\right]e^{i[\mathbf{x}\cdot\mathbf{v}(t)+\gamma(t)+\frac{\omega t}{2}]}.$$

Then, the spatial derivative with respect to x_j, for every $j = 1,\ldots,n$:

$$\frac{\partial}{\partial x_j}\psi(t,\mathbf{x}) = \left[\frac{\partial}{\partial y_j}u(\mathbf{y}) + i\mathbf{v}_j(t)u(\mathbf{y})\right]e^{i[\mathbf{x}\cdot\mathbf{v}(t)+\gamma(t)+\frac{\omega t}{2}]}.$$

In the end, the second spatial derivative with respect to x_j, for every $j = 1,\ldots,n$:

$$\frac{\partial^2}{\partial x_j^2}\psi(t,\mathbf{x}) = \left[\frac{\partial^2}{\partial y_j^2}u(\mathbf{y}) + 2i\mathbf{v}_j(t)\frac{\partial}{\partial x_j}u(\mathbf{y}) - |\mathbf{v}_j(t)|^2u(\mathbf{y})\right]e^{i[\mathbf{x}\cdot\mathbf{v}(t)+\gamma(t)+\frac{\omega t}{2}]}.$$

Now, we can plug everything inside equation (3.1) and get:

$$\begin{aligned} & i\hbar\left[-\dot{\mathbf{x}}\cdot\nabla_y u(\mathbf{y}) + iu(\mathbf{y})\left(\dot{\mathbf{v}}(t)\cdot\mathbf{y} + \dot{\mathbf{v}}(t)\cdot\mathbf{x}(t) + \dot{\gamma}(t) + \frac{\omega}{2}\right)\right] \\ & = -\frac{\hbar^2}{2m}\left[\Delta_{\mathbf{y}}u(\mathbf{y}) + 2i\mathbf{v}(t)\cdot\nabla u(\mathbf{y}) - |\mathbf{v}(t)|^2u(\mathbf{y})\right] \\ + & \quad \frac{1}{2}m\omega^2\left[|\mathbf{y}|^2 + 2\mathbf{x}(t)\cdot\mathbf{y} + |\mathbf{x}(t)|^2\right]u(\mathbf{y}), \end{aligned}$$

since the coefficient $e^{i[\mathbf{x}\cdot\mathbf{v}(t)+\gamma(t)+\frac{\omega t}{2}]}$ is never zero and appears in every term. Now, if we impose that the *profile* $u(\mathbf{x})$ satisfies the equation

$$-\frac{\hbar^2}{2m}\Delta u(\mathbf{x}) + \frac{1}{2}m\omega^2|\mathbf{x}|^2u(\mathbf{x}) + \frac{\hbar\omega}{2}u(\mathbf{x}) = 0, \quad \mathbf{x}\in\mathbb{R}^n,$$

we are left with the identity:

$$\begin{aligned}&i\hbar\left[-\dot{\mathbf{x}}\cdot\nabla_y u(\mathbf{y})+iu(\mathbf{y})\dot{\mathbf{v}}(t)\cdot(\mathbf{y}+\mathbf{x}(t)+\dot{\gamma}(t))\right]\\&=-\frac{\hbar^2}{2m}\left[+2i\mathbf{v}(t)\cdot\nabla u(\mathbf{y})-|\mathbf{v}(t)|^2u(\mathbf{y})\right]+\frac{1}{2}m\omega^2\left[2\mathbf{x}(t)\cdot\mathbf{y}+|\mathbf{x}(t)|^2\right]u(\mathbf{y}).\end{aligned}$$

Now, equating term by term, we obtain $-i\hbar\dot{\mathbf{x}}\cdot\nabla_y u(\mathbf{y})=-\frac{\hbar^2}{m}i\mathbf{v}(t)\cdot\nabla_y u(\mathbf{y})$ and hence

$$\dot{\mathbf{x}}(t)=\frac{\hbar}{m}\mathbf{v}(t).$$

Moreover, we get $-\hbar\dot{\mathbf{v}}(t)\cdot\mathbf{y}u(\mathbf{y})=m\omega^2\mathbf{x}(t)\cdot\mathbf{y}u(\mathbf{y})$ and so

$$\dot{\mathbf{v}}(t)u(\mathbf{y})=-\frac{m}{\hbar}\omega^2\mathbf{x}(t).$$

Now the remaining condition is

$$-\hbar\left(\dot{\mathbf{v}}(t)\cdot\mathbf{x}(t)+\dot{\gamma}(t)\right)=\frac{\hbar^2}{2m}|\mathbf{v}(t)|^2u(\mathbf{y})+\frac{1}{2}m\omega^2|\mathbf{x}(t)|^2u(\mathbf{y}),$$

which, using the other conditions, becomes

$$-\dot{\gamma}(t)=\frac{m\omega^2}{2\hbar}|\mathbf{x}(t)|^2-\frac{m}{2\hbar}|\dot{\mathbf{x}}(t)|^2$$

and so

$$\dot{\gamma}(t)=\frac{1}{\hbar}\mathcal{L}(\mathbf{x}(t),\dot{\mathbf{x}}(t);m,\omega),$$

with $\mathcal{L}(\mathbf{x}(t),\dot{\mathbf{x}}(t);m,\omega)=\frac{1}{2}m|\dot{\mathbf{x}}(t)|^2-\frac{1}{2}m\omega^2|\mathbf{x}(t)|^2$. This concludes the proof of the theorem. □

Remark 3.2.9. *The system*

$$\begin{cases}\dot{\mathbf{x}}(t)=\frac{\hbar}{m}\mathbf{v}(t)\\\dot{\mathbf{v}}(t)=-\frac{m}{\hbar}\omega^2\mathbf{x}(t)\end{cases}$$

can be reduced to the second order ODE:

$$\ddot{\mathbf{x}}+\omega^2\mathbf{x}=0,\tag{3.5}$$

which has the explicit solutions

$$\mathbf{x}(t)=\mathbf{x}_0\cos(\omega t)+\frac{\hbar}{m}\frac{\mathbf{v}_0}{\omega}\sin(\omega t),\tag{3.6}$$

where $\mathbf{x}_0$ *is the* initial position*,* $\mathbf{v}_0$ *is the* initial velocity*,* m *the* mass*,* $\hbar$ *the* Planck Constant *and* ω *is the* frequency*.*

3.2.2.2 Existence of Solitons and Multisolitons for the*Nonlinear Schrödinger Equation*

In this subsubsection, we present the most complete results about the existence and properties of a single soliton solutions for the *Nonlinear Schrödinger Equation*, mainly due to Berestycki-Lions [11], Berestycki-Lions-Peletier [12], Gidas-Ni-Nirenberg [56] and Kwong [75]. Moreover, we present a theorem by Martel and Merle [88] which proves the existence of Multi-solitons solutions for such equation. We will not present the full proof here, because it is involved, and we refer to [88] for more details. We consider the following *Nonlinear Schrödinger Equation*

$$\begin{cases} i\frac{\partial}{\partial t}\psi(t,x) = -\Delta_x\psi(t,x) - |\psi(t,x)|^{p-1}\psi(t,x), \quad (t,x) \in \mathbb{R}\times\mathbb{R}^n \\ \psi(0,x) = \psi_0(x), \end{cases} \tag{3.7}$$

where $n \geq 1$, $1 < p < 1 + \frac{4}{n}$ is the L^2-*subcritical exponent*. For this equation, there are soliton solutions, which share very similar property with the solutions of (3.1). These solutions are solitary waves of the form $\psi(t,x) = e^{i\omega t}Q_\omega(x)$ for some $\omega > 0$ and where $Q_\omega \in H^1(\mathbb{R}^n)$ is a solution of

$$\Delta Q_\omega + Q_\omega^p = \omega Q_\omega, \quad Q_\omega > 0. \tag{3.8}$$

These solutions Q_ω can be computed explicitly in dimension $n = 1$ and take the form

$$Q_\omega(x) = \omega^{\frac{1}{p-1}}\left(\frac{p+1}{2\cosh^2\left(\frac{p-1}{2}\omega^{\frac{1}{2}}x\right)}\right)^{p-1},$$

up to symmetries. In any dimension $n \geq 1$, they are radially symmetric, exponentially decaying, unique up to symmetries and minimize the so called *Energy Functional*. We refer to [11], [12], [56] and [75] for more details. By the symmetries of equation (3.7), for any $v_0 \in \mathbb{R}^n$, $x_0 \in \mathbb{R}^n$ and $\gamma_0 \in \mathbb{R}$

$$\psi(t,x) = Q_\omega(x - x_0 - v_0 t)e^{i[\frac{1}{2}v_0\cdot x - \frac{t}{4}|v_0|^2 + \omega t + \gamma_0]} \tag{3.9}$$

is still a solution of (3.7) and represents a soliton moving on the line $x = x_0 + v_0 t$. In this framework, we can state a theorem about the existence of multi-solitary waves for the *Subcritical NLS* due to Martel-Merle [88].

Theorem 3.2.10. *[88] Let $1 < p < 1 + 4/n$. Let $K \in \mathbb{N}^*$. For any $k \in \{1, \ldots, K\}$, let $\omega_k > 0$, $v_k \in \mathbb{R}^n$ and $\gamma_k^0 \in \mathbb{R}$. Assume that for any $k \neq k'$, we have $v_k \neq v_{k'}$. Let*

$$R_k(t,x) = Q_{\omega_k}(x - x_k^0 - v_k^0 t) e^{i[\frac{1}{2} v_k^0 \cdot x - \frac{t}{4}|v_k^0|^2 + \omega_k t + \gamma_k^0]}.$$

Then, there exists an $H^1(\mathbb{R}^n)$ solution $U(t)$ of (3.7) such that, for all $t \geq 0$, we have:

$$\left\| U(t) - \Sigma_{k=1}^K R_k(t) \right\|_{H^1(\mathbb{R}^n)} \leq C e^{-\theta_0 t}$$

for some $\theta_0 > 0$ and $C > 0$.

3.2.2.3 Measure of Amalgamation

In the following, we consider the problem of amalgamation in the context of *Partial Differential Equations* (PDEs). We use the following *measure of amalgamation.*

Definition 3.2.11. *Consider two solutions $\psi(t, \mathbf{x})$ and $\phi(t, \mathbf{x})$ of equation (3.1). The* Measure of Correlation *between $\psi(t, \mathbf{x})$ and $\phi(t, \mathbf{x})$ is given by*

$$Corr(\psi_i(t,x), \psi_j(t,x)) = \frac{Cov(\psi_i(t,x), \psi_j(t,x))}{Var(\psi_i(t,x))^{\frac{1}{2}} Var(\psi_j(t,x))^{\frac{1}{2}}}, \tag{3.10}$$

where

$$Cov(\psi_i(t,\mathbf{x}), \psi_j(t,\mathbf{x})) := \Re \int_{\mathbb{R}^n} (\mathbf{x} - \mu_i)(\mathbf{x} - \mu_j) \psi_i(t,\mathbf{x}) \bar{\psi}_i(t,\mathbf{x}) d\mathbf{x} \tag{3.11}$$

Remark 3.2.12. *We will actually use mainly the* Covariance $Cov(\psi_i(t,\mathbf{x}), \psi_j(t,\mathbf{x}))$, *instead of the* Measure of Correlation $Corr(\psi_i(t,x), \psi_j(t,x))$ *for the following reasons.*

First of all, here it is not important to have a measure of homogeneity 0 *(see [59]), because (similarly as in the two body problem) the bigger the mass of the solutions is, the more the attraction between the solutions increases and so it makes sense to take this into account in our* Measure of Amalgamation.

Second, since we chose the constants in order to make the solitons all of mass one, we could simply divide $Cov(\psi_i(t,\mathbf{x}), \psi_j(t,\mathbf{x}))$ by the masses of each solutions to get something *of homogeneity zero (be careful that this would not be equal to $Corr(\psi_i(t,x), \psi_j(t,x))$). Third, this choice makes the computations less obscure, since they are already involved.*

This *Measure of Amalgamation* seems different from the discussed in [59] and from the structure outlined above. Here we see how, in fact, this measure is not that dissimilar. Consider two solitons $\psi_i = \rho_i e^{i\theta_i} = \Re(\psi_i) + iIm(\psi_i)$, $i = 1, 2$. Then, build the following contingency table

$$\mathbf{a} = [\rho_1 \cos(\theta_1), \rho_1 \sin(\theta_1); \rho_2 \cos(\theta_2), \rho_2 \sin(\theta_2)].$$

Take as measure of amalgamation $\alpha(\mathbf{a}) = det(\mathbf{a})$, so $\alpha(\mathbf{a}) = \rho_1\rho_2 \sin(\theta_1 - \theta_2)$. Now since there is an in some sense "hidden variable x", we integrate over it and get the correlation $Cov(\psi_i(t, \mathbf{x}), \psi_j(t, \mathbf{x}))$, apart from a translation in the phase of $\frac{\pi}{2}$. This small adjustment in the phase is due to the fact that we want to have the highest amalgamation when the solitons are multiples one of the other and minimal when the phase is opposite.

Remark 3.2.13. *One could argue that a paradoxical situation appears also when*

$$Cov(\psi_1(t, \mathbf{x}), \psi_2(t, \mathbf{x})) << -1.$$

and

$$0 \leq Cov(\psi_3(t, \mathbf{x}), \psi_4(t, \mathbf{x})) << 1$$

but

$$Cov(\psi_1(t, \mathbf{x}) + \psi_3(t, \mathbf{x}), \psi_2(t, \mathbf{x}) + \psi_4(t, \mathbf{x})) \geq 0.$$

This makes perfect sense, but it is not how the Simpson's Paradox *has been defined in the literature. See for example [94].*

3.2.3 Proof of Theorem 3.2.1

In this subsection, we give the proof of Theorem 3.2.1. We will divide it in different subsubsections accordingly to the space dimension and if the case is stationary or non-stationary.

3.2.3.1 The case $n = 1$

In this subsubsection, we consider the case of spatial dimension $n = 1$.

Remark 3.2.14. *In spatial dimension $n = 1$, the solutions of equation (3.3) can be computed explicitly. In fact, just by plugging in the equation the* ansatz, *the solutions can* *be verified to be*

$$u(x) = e^{-\frac{m\omega}{2\hbar}x^2}.$$

Since equation (3.1) and (3.3) are linear equations, then every multiple of a solution is still a solution. For normalizing reasons, in the following we will consider the following solution

$$u(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{m\omega}{2\hbar}x^2}.$$

3.2.3.2 The stationary case

In this subsubsection, we will always assume $x(t) = x_i$ with $x_i \in \mathbb{R}$ independent of time for any $i = 1, \dots, 4$. We recall the definition of *Measure of Correlation* that we will use.

Definition 3.2.15. *Consider two solutions $\psi(t, \mathbf{x})$ and $\phi(t, \mathbf{x})$ of equation (3.1). The* Measure of Correlation *between $\psi(t, \mathbf{x})$ and $\phi(t, \mathbf{x})$ is given by*

$$Corr(\psi_i(t, \mathbf{x}), \psi_j(t, \mathbf{x})) = \frac{Cov(\psi_i(t, \mathbf{x}), \psi_j(t, \mathbf{x}))}{Var(\psi_i(t, \mathbf{x}))^{\frac{1}{2}} Var(\psi_j(t, \mathbf{x}))^{\frac{1}{2}}},$$

where

$$Cov(\psi_i(t, \mathbf{x}), \psi_j(t, \mathbf{x})) := \Re \int_{\mathbb{R}^n} (\mathbf{x} - \mu_i)(\mathbf{x} - \mu_j)\psi_i(t, \mathbf{x})\overline{\psi}_i(t, \mathbf{x}) d\mathbf{x}$$

Now, we are ready to compute the *Measure of Correlation* for two solitons in spatial dimension $n = 1$ and in the stationary case.

Proposition 3.2.16. *Consider two* moving solitons*:*

$$\psi_i(t, x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x \cdot v_i(t) + \gamma_i(t) + \frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x - x_i(t)|^2},$$

and

$$\psi_j(t, x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x \cdot v_j(t) + \gamma_j(t) + \frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x - x_j(t)|^2},$$

for $1 \leq i \leq j \leq 4$ and with $\mathbf{x}(t)$, $\mathbf{v}(t)$ and $\gamma(t)$ as in Theorem 3.2.8. Consider the case in which, for every $t \in \mathbb{R}$, one has that $x_k(t) = x_k$, for every $k = 1, \dots, N$ independent of time. Then, the Covariance *between these two solitons is given by:*

$$Cov(\psi_i(t, x), \psi_j(t, x)) = \frac{1}{2}\cos(\gamma_i - \gamma_j)\left[\frac{\hbar}{m\omega} - \frac{1}{2}|x_i - x_j|^2\right] e^{-\frac{m\omega}{4\hbar}|x_i - x_j|^2} \quad (3.12)$$

Proof. We need to use the definition of Covariance and the shape of the solitons.

$$
\begin{aligned}
&Cov(\psi_i(t,x),\psi_j(t,x)) = \Re \int_{\mathbb{R}^n} (\mathbf{x}-\mu_i)(\mathbf{x}-\mu_j)\psi_i(t,\mathbf{x})\overline{\psi}_i(t,\mathbf{x})d\mathbf{x} \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \Re \int_{\mathbb{R}} (x-\mu_i)(x-\mu_j) e^{i[\gamma_i + \frac{\omega t}{2} - \gamma_j - \frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x-x_i(t)|^2 - \frac{m\omega}{2\hbar}|x-x_j(t)|^2} dx \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \Re e^{i[\gamma_i-\gamma_j]} \int_{\mathbb{R}} (x-\mu_i)(x-\mu_j) e^{-\frac{m\omega}{2\hbar}[2x^2 - 2x(x_i+x_j) + x_i^2 + x_j^2]} dx \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{2\hbar}[x_i^2+x_j^2]} \Re e^{i[\gamma_i-\gamma_j]} \times \\
&\times \int_{\mathbb{R}} (x-\mu_i)(x-\mu_j) e^{-\frac{m\omega}{\hbar}[x^2 - x(x_i+x_j) + \frac{1}{4}(x_i+x_j)^2] + \frac{m\omega}{4\hbar}(x_i+x_j)^2} dx \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{2\hbar}[x_i^2+x_j^2] + \frac{m\omega}{4\hbar}(x_i+x_j)^2} \cos(\gamma_i-\gamma_j) \\
\times \quad &\int_{\mathbb{R}} (x-\mu_i)(x-\mu_j) e^{-\frac{m\omega}{\hbar}\left|x - \frac{1}{2}(x_i+x_j)\right|^2} dx.
\end{aligned}
$$

Now, by changing variables to $y = x - \frac{x_i+x_j}{2}$ and using the precise value of μ_i (See Lemma 3.2.21 below), we get:

$$
\begin{aligned}
&Cov(\psi_i(t,x),\psi_j(t,x)) = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}|x_i-x_j|^2} \cos(\gamma_i-\gamma_j) \\
\times \quad &\int_{\mathbb{R}} \left(y + \frac{x_i+x_j}{2} - x_i\right)\left(y + \frac{x_i+x_j}{2} - x_j\right) e^{-\frac{m\omega}{\hbar}|y|^2} dy \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}|x_i-x_j|^2} \cos(\gamma_i-\gamma_j) \int_{\mathbb{R}} \left(y^2 - \frac{(x_i-x_j)^2}{4}\right) e^{-\frac{m\omega}{\hbar}|y|^2} dy \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}|x_i-x_j|^2} \cos(\gamma_i-\gamma_j)\,(I_1 - I_2)
\end{aligned}
$$

with

$$I_1 = \int_{\mathbb{R}} y^2 e^{-\frac{m\omega}{\hbar}y^2} dy.$$

and

$$I_2 = \int_{\mathbb{R}} \frac{(x_i-x_j)^2}{4} e^{-\frac{m\omega}{\hbar}y^2} dy.$$

In both I_1 and I_2 we use the change of variables $z = y\left(\frac{m\omega}{\pi\hbar}\right)^{1/2}$, which implies $y =$

$\left(\frac{m\omega}{\pi\hbar}\right)^{-1/2} z$ and so $dy = \left(\frac{m\omega}{\pi\hbar}\right)^{-1/2} dz$. This leads to

$$I_1 = \int_{\mathbb{R}} y^2 e^{-\frac{m\omega}{\hbar}y^2} dy = \left(\frac{m\omega}{\hbar}\right)^{-1} \int_{\mathbb{R}} z^2 e^{-z^2} \left(\frac{m\omega}{\hbar}\right)^{-1/2} dz = \left(\frac{m\omega}{\hbar}\right)^{-3/2} \frac{\pi^{1/2}}{2}$$

and to

$$\begin{aligned} I_2 &= \int_{\mathbb{R}} \frac{(x_i - x_j)^2}{4} e^{-\frac{m\omega}{\hbar}y^2} dy \\ &= \left(\frac{m\omega}{\hbar}\right)^{-1/2} \frac{(x_i - x_j)^2}{4} \int_{\mathbb{R}} e^{-z^2} dz = \left(\frac{m\omega}{\hbar}\right)^{-1/2} \frac{(x_i - x_j)^2}{4} \pi^{\frac{1}{2}}. \end{aligned}$$

Now, putting everything together, we get

$$\begin{aligned} &Cov(\psi_i(t,x), \psi_j(t,x)) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}|x_i - x_j|^2} \cos(\gamma_i - \gamma_j)\,(I_1 - I_2) \\ &= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}|x_i - x_j|^2} \cos(\gamma_i - \gamma_j) \left(\left(\frac{m\omega}{\hbar}\right)^{-3/2} \frac{\pi^{1/2}}{2} - \frac{(x_i - x_j)^2}{4} \left(\frac{m\omega}{\hbar}\right)^{-1/2} \pi^{\frac{1}{2}}\right) \\ &= \frac{1}{2}\cos(\gamma_i - \gamma_j) \left[\frac{\hbar}{m\omega} - \frac{1}{2}|x_i - x_j|^2\right] e^{-\frac{m\omega}{4\hbar}|x_i - x_j|^2}. \end{aligned}$$

This completes the proof. □

Corollary 3.2.17. *Consider two* moving solitons*:*

$$\psi_i(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_i(t) + \gamma_i(t) + \frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x - x_i(t)|^2},$$

and

$$\psi_j(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_j(t) + \gamma_j(t) + \frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x - x_j(t)|^2},$$

for $1 \le i \le j \le 4$ and with $\mathbf{x}(t)$, $\mathbf{v}(t)$ and $\gamma(t)$ as in Theorem 3.2.8. Consider the case in which, for every $t \in \mathbb{R}$, one has that $x_k(t) = x_k$, for every $k = 1, \ldots, 4$ independent of time. Then, the Measure of Correlation *between these two solitons is given by:*

$$Corr(\psi_i(t,x), \psi_j(t,x)) = \cos(\gamma_i - \gamma_j) \left[1 - \frac{m\omega}{2\hbar}|x_i - x_j|^2\right] e^{-\frac{m\omega}{4\hbar}|x_i - x_j|^2}.$$

Proof. Since

$$Cov(\psi_i(t,x), \psi_j(t,x)) = \frac{1}{2}\cos(\gamma_i - \gamma_j) \left[\frac{\hbar}{m\omega} - \frac{1}{2}|x_i - x_j|^2\right] e^{-\frac{m\omega}{4\hbar}|x_i - x_j|^2}$$

and $Var(\psi_i(t,x)) = Var(\psi_j(t,x)) = \frac{\hbar}{2m\omega}$ (See Lemma 3.2.22 below), then, by definition of *Measure of Correlation*, one gets

$$\begin{aligned} Corr(\psi_i(t,x), \psi_j(t,x)) &= \frac{Cov(\psi_i(t,x), \psi_j(t,x))}{Var(\psi_i(t,x))^{\frac{1}{2}} Var(\psi_j(t,x))^{\frac{1}{2}}} \\ &= \cos(\gamma_i - \gamma_j) \left[1 - \frac{m\omega}{2\hbar}|x_i - x_j|^2\right] e^{-\frac{m\omega}{4\hbar}|x_i - x_j|^2}. \end{aligned}$$

This completes the proof of the corollary. □

Now, we are ready to show that, with a proper choice of the parameters, the *Simpson's Paradox* occurs.

Proof of Theorem 3.2.1. The proof consists in finding parameters such that the *Simpson's Reversal* occurs, namely such that

$$Cov(\psi_1(t,x), \psi_2(t,x)) > 0,$$

$$Cov(\psi_3(t,x), \psi_4(t,x)) > 0$$

but

$$Cov(\psi_1(t,x) + \psi_3(t,x), \psi_2(t,x) + \psi_4(t,x)) < 0$$

or vice versa,

$$Cov(\psi_1(t,x), \psi_2(t,x)) < 0,$$

$$Cov(\psi_3(t,x), \psi_4(t,x)) < 0$$

but

$$Cov(\psi_1(t,x) + \psi_3(t,x), \psi_2(t,x) + \psi_4(t,x)) > 0.$$

Now, we define

$$L_{ij}^2 := \frac{m\omega}{2\hbar}|x_i - x_j|^2$$

so that $Cov(\psi_i(t,x), \psi_j(t,x))$ can be rewritten in the following way:

$$Cov(\psi_i(t,x), \psi_j(t,x)) = \frac{\hbar}{2m\omega} \cos(\gamma_i - \gamma_j) \left[1 - L_{ij}^2\right] e^{-\frac{1}{2}L_{ij}^2}.$$

First, we treat the case $\gamma_i = \gamma_j$, for every $i, j = 1, \ldots, 4$. We can restate our hypotheses and thesis in the following way: we suppose that $0 < L_{12} < 1$ and $0 < L_{34} < 1$ and we have to prove that there exists an admissible choice of $0 < L_{12} < 1$ and $0 < L_{34} < 1$, L_{23}

and L_{14} such that

$$\left[1-L_{12}^2\right]e^{-\frac{1}{2}L_{12}^2}+\left[1-L_{23}^2\right]e^{-\frac{1}{2}L_{23}^2}+\left[1-L_{34}^2\right]e^{-\frac{1}{2}L_{34}^2}+\left[1-L_{14}^2\right]e^{-\frac{1}{2}L_{14}^2}<0.$$

Since we are in dimension $n=1$, we can choose $x_1<x_2<x_3<x_4$. This implies that $L_{14}=L_{12}+L_{23}+L_{34}$ and so that we have to find an admissible choice of $0<L_{12}<1$ and $0<L_{34}<1$ and L_{23} such that

$$\begin{aligned}&\left[1-L_{12}^2\right]e^{-\frac{1}{2}L_{12}^2}+\left[1-L_{23}^2\right]e^{-\frac{1}{2}L_{23}^2}+\\&+\left[1-L_{34}^2\right]e^{-\frac{1}{2}L_{34}^2}+\left[1-(L_{12}+L_{23}+L_{34})^2\right]e^{-\frac{1}{2}(L_{12}+L_{23}+L_{34})^2}<0.\end{aligned}$$

We choose $L_{12}^2=1-\epsilon_1^2$, so that $Cov(\psi_1(t,x),\psi_2(t,x))>0$,, $L_{34}^2=1-\epsilon_2^2$ so that $Cov(\psi_3(t,x),\psi_4(t,x))>0$, and $L_{23}^2=1+\delta^2$ with $\epsilon_1<<1$, $\epsilon_2<<1$. Therefore, we get:

$$\begin{aligned}&\epsilon_1^2e^{-\frac{1}{2}(1-\epsilon_1^2)}-\delta^2e^{-\frac{1}{2}(1+\delta^2)}+\epsilon_2^2e^{-\frac{1}{2}(1-\epsilon_2^2)}+\\&+\left[1-\left(\sqrt{1-\epsilon_1^2}+\sqrt{1-\epsilon_2^2}+\sqrt{1+\delta^2}\right)^2\right]e^{-\frac{1}{2}\left(\sqrt{1-\epsilon_1^2}+\sqrt{1-\epsilon_2^2}+\sqrt{1+\delta^2}\right)^2}<0.\end{aligned}$$

Now, notice that $\left(\sqrt{1-\epsilon_1^2}+\sqrt{1-\epsilon_2^2}+\sqrt{1+\delta^2}\right)^2>1$ and so that

$$\left[1-(\sqrt{1-\epsilon_1^2}+\sqrt{1-\epsilon_2^2}+\sqrt{1+\delta^2})^2\right]e^{-\frac{1}{2}\left(\sqrt{1-\epsilon_1^2}+\sqrt{1-\epsilon_2^2}+\sqrt{1+\delta^2}\right)^2}<0.$$

This implies that we just need to find ϵ_1, ϵ_2 and δ such that

$$\epsilon_1^2e^{-\frac{1}{2}(1-\epsilon_1^2)}-\delta^2e^{-\frac{1}{2}(1+\delta^2)}+\epsilon_2^2e^{-\frac{1}{2}(1-\epsilon_2^2)}<0$$

to have that *Simpson's Paradox* occurs. We choose $\epsilon_1=\epsilon_2=\alpha>0$ and $\delta=k\alpha>0$ and so we just need to find k and α such that

$$2\alpha^2e^{-\frac{1}{2}(1-\alpha^2)}-k^2\alpha^2e^{-\frac{1}{2}(1+(k\alpha)^2)}<0.$$

Reorganizing all the terms and taking the logarithm to both sides of the inequality, one finds that, if one chooses k and α such that

$$\alpha^2<\frac{2\log\left(\frac{k^2}{2}\right)}{k^2+1},$$

then the *Simpson's Paradox* occurs. It is enough to choose for example $\alpha = 10^{-100}$ and $k = 100$ and both $\alpha^2 < \frac{2\log\left(\frac{k^2}{2}\right)}{k^2+1}$ and $0 < k\alpha < 1$ are satisfied. This concludes the proof of the case $\gamma_i = \gamma_j$, for every $i, j = 1, \dots, 4$ in spatial dimension $n = 1$.
Now, we pass to the case $\gamma_i = \gamma_j$ if and only if $i = j$, $i, j = 1, \dots, 4$.

Recall that we defined

$$L_{ij} := \frac{m\omega}{2\hbar}|x_i - x_j|^2$$

and so $Cov(\psi_i(t,x), \psi_j(t,x))$ can be rewritten in the following way:

$$Cov(\psi_i(t,x), \psi_j(t,x)) = \frac{\hbar}{2m\omega}\cos(\gamma_i - \gamma_j)\left[1 - L_{ij}^2\right]e^{-\frac{1}{2}L_{ij}^2}.$$

for every $i, j = 1, \dots, 4$. We assume that

$$Cov(\psi_1(t,x), \psi_2(t,x)) > 0,$$

and that

$$Cov(\psi_3(t,x), \psi_4(t,x)) > 0$$

and want to prove that it can happen that

$$Cov(\psi_1(t,x) + \psi_3(t,x), \psi_2(t,x) + \psi_4(t,x)) < 0.$$

We choose $\gamma_2 - \gamma_1 = \frac{\pi}{2} - \epsilon$ and $\gamma_4 - \gamma_3 = \frac{\pi}{2} - \epsilon$ with $0 < \epsilon << 1$. With this we get:

$$\begin{aligned}
&\frac{2m\omega}{\hbar}Cov(\psi_1(t,x) + \psi_3(t,x), \psi_2(t,x) + \psi_4(t,x)) = \\
&\cos(\gamma_2 - \gamma_1)[1 - L_{12}^2]e^{-\frac{1}{2}L_{12}^2} + \cos(\gamma_3 - \gamma_2)[1 - L_{23}^2]e^{-\frac{1}{2}L_{23}^2} \\
&+ \cos(\gamma_4 - \gamma_3)[1 - L_{34}^2]e^{-\frac{1}{2}L_{34}^2} + \cos(\gamma_1 - \gamma_4)[1 - L_{41}^2]e^{-\frac{1}{2}L_{41}^2} = \\
&\sin(\epsilon)[1 - L_{12}^2]e^{-\frac{1}{2}L_{12}^2} + \cos(\gamma_3 - \gamma_2)[1 - L_{23}^2]e^{-\frac{1}{2}L_{23}^2} \\
&+ \sin(\epsilon)[1 - L_{34}^2]e^{-\frac{1}{2}L_{34}^2} + \cos(\gamma_1 - \gamma_4)[1 - L_{41}^2]e^{-\frac{1}{2}L_{41}^2}.
\end{aligned}$$

Now, we choose $\gamma_3 - \gamma_2 = \pi - \delta$ and so this implies that

$$\gamma_4 - \gamma_1 = \gamma_4 - \gamma_3 + \gamma_3 - \gamma_2 + \gamma_2 - \gamma_1 = \frac{\pi}{2} - \epsilon + \pi - \delta + \frac{\pi}{2} - \epsilon = 2\pi - (2\epsilon + \delta).$$

Therefore, we obtain:

$$\begin{aligned}
&\frac{2m\omega}{\hbar}Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x)) = \\
&\sin(\epsilon)[1-L_{12}^2]e^{-\frac{1}{2}L_{12}^2}+\cos(\gamma_3-\gamma_2)[1-L_{23}^2]e^{-\frac{1}{2}L_{23}^2} \\
&+\sin(\epsilon)[1-L_{34}^2]e^{-\frac{1}{2}L_{34}^2}+\cos(\gamma_1-\gamma_4)[1-L_{41}^2]e^{-\frac{1}{2}L_{41}^2} = \\
&\sin(\epsilon)[1-L_{12}^2]e^{-\frac{1}{2}L_{12}^2}+\cos(\pi-\delta)[1-L_{23}^2]e^{-\frac{1}{2}L_{23}^2} \\
&+\sin(\epsilon)[1-L_{34}^2]e^{-\frac{1}{2}L_{34}^2}+\cos(2\epsilon+\delta)[1-L_{41}^2]e^{-\frac{1}{2}L_{41}^2}.
\end{aligned}$$

Now, we can choose $L_{12}=L_{23}=L_{34}=L<1$ (which implies $L_{14}=3L$), so that we get

$$\begin{aligned}
&\frac{2m\omega}{\hbar}Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x)) = \\
&\sin(\epsilon)[1-L_{12}^2]e^{-\frac{1}{2}L_{12}^2}+\cos(\pi-\delta)[1-L_{23}^2]e^{-\frac{1}{2}L_{23}^2} \\
&+\sin(\epsilon)[1-L_{34}^2]e^{-\frac{1}{2}L_{34}^2}-\cos(2\epsilon+\delta)[1-L_{41}^2]e^{-\frac{1}{2}L_{41}^2} = \\
&2\sin(\epsilon)[1-L^2]e^{-\frac{1}{2}L^2}-\cos(\delta)[1-L^2]e^{-2L^2}+\cos(2\epsilon+\delta)[1-9L^2]e^{-\frac{9}{2}L^2}.
\end{aligned}$$

for $\epsilon>0$. So, if you choose ϵ and δ small enough and $\frac{1}{3}<L<1$ but $L\simeq 1$, we get

$$\begin{aligned}
&\frac{2m\omega}{\hbar}Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x)) = \\
&2\sin(\epsilon)[1-L^2]e^{-\frac{1}{2}L^2}-\cos(\delta)[1-L^2]e^{-2L^2}+\cos(2\epsilon+\delta)[1-9L^2]e^{-\frac{9}{2}L^2} \\
\simeq\ &-8e^{-9/2}<0.
\end{aligned}$$

and so we get

$$Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x))<0,$$

This completes the proof of the case $\gamma_i=\gamma_j$ if and only if $i=j$, $i,j=1,\ldots,4$. So the proof of the stationary case in dimension $n=1$ is complete. □

Remark 3.2.18. *In the case of the contingency tables, one encounters matrices of the form*

$$A_i=[a_i,b_i;c_i,d_i],$$

with $i=1,2$. Using the notation of these chapters, this means that "$Cov(A^i,A^j)=\sin(\gamma_i-\gamma_j)\rho_i\rho_j$" if you rewrite A_i in polar form: $a_i=\rho_i\cos(\gamma_i)$, $b_i=\rho_i\sin(\gamma_i)$, $c_i=$

$a_j = \rho_j cos(\gamma_j)$ *and* $d_i = b_j = \rho_j \sin(\gamma_j)$. *So, we have a* Measure of Amalgamation *of the form* $Cov(\psi_1(t,x), \psi_2(t,x)) = R(\rho_i, \rho_j)\Theta(\gamma_i - \gamma_j)$ *(a radial part times an angular part). This structure is very similar to that of the* Quantum Harmonic Oscillator. *The difference is that, in the case of the* Quantum Harmonic Oscillator, *one has* $R(\rho_i, \rho_j) = |\rho_i - \rho_j|^2 e^{-\frac{|\rho_i - \rho_j|^2}{2}}$, *not always increasing in* ρ_i, ρ_j *and this is due to the tail interaction between the quantum particles, which is absent in the usual contingency table case where we get* $R(\rho_i, \rho_j) = \rho_i \rho_j$.

Remark 3.2.19. *The case when* $L_{ij} > 1$, *namely when*

$$\hbar < \frac{m\omega}{2}|x_i - x_j|^2$$

is the Semiclassical Regime, *namely the regime where the* Planck Constant $\hbar$ *is smaller the all the other physical quantities present. We call the case when* $L_{ij} = 1$, *namely when*

$$\hbar = \frac{m\omega}{2}|x_i - x_j|^2$$

Uncorrelation Regime, *since in this case, due to a non-trivial interplay between the physical quantities one has uncorrelation of the solitons (*$Cov(\psi_i(t,x), \psi_j(t,x))$*=0). When* $L_{ij} < 1$, *namely when*

$$\hbar > \frac{m\omega}{2}|x_i - x_j|^2$$

we will talk about Anti-Semiclassical Regime, *because now certain physical constants are even smaller than the* Planck Constant $\hbar$. *We underline that, in the case* $\gamma_i = \gamma_j$ *for any* $i, j = 1, \dots, n$, *the* Quantum Effect *becomes evident because the angles between the solitons do not play a role, as in the classical (usual contingency tables regime) case, but it is the presence of both the* Semiclassical Regime *and the* Anti-Semiclassical Regime *that cause the* Simpson's Paradox.

Remark 3.2.20. *One can read the* Simpson's Paradox *as the following fact. Even if the soliton* ψ_1 *repels* ψ_3 *(*$Cov(\psi_i(t,x), \psi_j(t,x)) < 0$*) and* ψ_2 *repels* ψ_4 *(*$Cov(\psi_i(t,x), \psi_j(t,x)) < 0$*), the multi-soliton* $\psi_1 + \psi_3$ *can attract* $\psi_2 + \psi_4$ *(*$Cov(\psi_i(t,x), \psi_j(t,x)) > 0$*). Similarly by reversing all the inequalities.*

3.2.3.3 The non-stationary case $n = 1$

Now, we pass to the non-stationary case in spatial dimension $n = 1$. Consider moving solitons of the form

$$\psi_i(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_i(t) + \gamma_i(t) + \frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x - x_i(t)|^2}, \tag{3.13}$$

for $i = 1, \dots, N$ with N the number of solitons and with $\mathbf{x}(t)$, $\mathbf{v}(t)$ and $\gamma(t)$ as in Theorem 3.2.8. For these objects, we compute how the *center of Mass* moves and how the *Variance* changes under the evolution of (3.1).

Lemma 3.2.21. *The* center of Mass *of each $\psi_i(t, x)$ as defined in (3.13) is given by*

$$\mu_i := \int_{\mathbb{R}} x|\psi(t,x)|^2 dx = x_i(t), \tag{3.14}$$

for every $i = 1, \dots, N$.

Proof. First, we recall the following fact about integrals involving *Gaussians*

$$\begin{aligned}\int_{\mathbb{R}} e^{-ax^2} dx &= \left(\int_{\mathbb{R}^2} e^{-a(x^2+y^2)} dxdy\right)^{\frac{1}{2}} = (2\pi)^{1/2} \left(\int_0^{+\infty} e^{-a\rho^2} \rho d\rho\right)^{\frac{1}{2}} \\ &= -\left(\frac{\pi}{a}\right)^{\frac{1}{2}} e^{-a\rho^2}\Big|_0^{+\infty} = \left(\frac{\pi}{a}\right)^{\frac{1}{2}},\end{aligned}$$

for every $a > 0$. Therefore, for every $i = 1, \dots, N$ and using the change of variables $y := x - x(t)$, we have

$$\begin{aligned}\mu_i &:= \int_{\mathbb{R}} x|\psi(t,x)|^2 dx = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{2}} \int_{\mathbb{R}} x e^{-\frac{m\omega}{\hbar}[x-x_i(t)]^2} dx \\ &= \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{2}} \int_{\mathbb{R}} [y + x_i(t)] e^{-\frac{m\omega}{\hbar}y^2} dy \\ &= \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{2}} \int_{\mathbb{R}} x(t) e^{-\frac{m\omega}{\hbar}y^2} dy = x_i(t).\end{aligned}$$

□

Lemma 3.2.22. *The* Variance *of each $\psi_i(t, x)$ as defined in (3.13) is given by*

$$Var(\psi_i(t,x)) := \int_{\mathbb{R}} |x|^2 |\psi(t,x)|^2 dx - \left(\int_{\mathbb{R}} x|\psi(t,x)|^2 dx\right)^2 = \frac{\hbar}{2m\omega}, \tag{3.15}$$

for every $i = 1, \dots, N$.

Proof. First of all we notice that

$$\left(\int_{\mathbb{R}} x|\psi(t,x)|^2 dx\right)^2 = |\mu_i|^2 = |x_i(t)|^2$$

and so we need to compute just $\int_{\mathbb{R}} |x|^2 |\psi(t,x)|^2 dx$.

$$
\begin{aligned}
Var(\psi_i(t,x)) &:= -|x_i(t)|^2 + \int_{\mathbb{R}} |x|^2 |\psi(t,x)|^2 dx \\
&= -|x_i(t)|^2 + \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \int_{\mathbb{R}} |x|^2 e^{-\frac{m\omega}{\hbar}[x-x_i(t)]^2} dx \\
&= -|x_i(t)|^2 + \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \int_{\mathbb{R}} [|y|^2 + 2x_i(t)y + |x_i(t)|^2] e^{-\frac{m\omega}{\hbar}y^2} dy \\
&= -|x_i(t)|^2 + \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \int_{\mathbb{R}} [|y|^2 + |x_i(t)|^2] e^{-\frac{m\omega}{\hbar}y^2} dy \\
&= (1-1)|x_i(t)|^2 + \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \int_{\mathbb{R}} |y|^2 e^{-\frac{m\omega}{\hbar}y^2} dy = \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \int_{\mathbb{R}} |y|^2 e^{-\frac{m\omega}{\hbar}y^2} dy \\
&= \left(\frac{m\omega}{\hbar}\right)^{1/2-1-1/2} \pi^{-1/2} \int_{\mathbb{R}} |z|^2 e^{-z^2} dz = \frac{\hbar}{2m\omega}.
\end{aligned}
$$

□

Remark 3.2.23. *These last two lemmas are in accordance with the fact that solitons are entities which travel linearly in space and do not spread spatially over time.*

Now, we can compute how the *Measure of Correlation* between two solitons vary over time.

Proposition 3.2.24. *Consider two* moving solitons*:*

$$
\psi_i(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_i(t)+\gamma_i(t)+\frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x-x_j(t)|^2},
$$

and

$$
\psi_j(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_j(t)+\gamma_j(t)+\frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x-x_j(t)|^2},
$$

for $1 \le i \le j \le N$ *and with* $\mathbf{x}(t)$*,* $\mathbf{v}(t)$ *and* $\gamma(t)$ *as in Theorem 3.2.8. Then, the* Covariance *between these two solitons is given by:*

$$
Cov(\psi_i(t,x),\psi_j(t,x)) = \quad e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \tag{3.16}
$$

$$
\times \frac{1}{2} \cos\left\{ i[\gamma_i(t) - \gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]\right\} \tag{3.17}
$$

$$
\times \left(\frac{\hbar}{m\omega} - \left(\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{2\omega^2} + \frac{1}{2}|x_i(t)-x_j(t)|^2 \right) \right). \tag{3.18}
$$

Proof. Again, we need to use the definition of Covariance and the shape of the solitons.

$$Cov(\psi_i(t,x),\psi_j(t,x)) = \Re\int_{\mathbb{R}^n}(\mathbf{x}-\mu_i)(\mathbf{x}-\mu_j)\psi_i(t,\mathbf{x})\overline{\psi}_i(t,\mathbf{x})d\mathbf{x}$$
$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2}\times$$
$$\times\Re\int_{\mathbb{R}}(x-\mu_i)(x-\mu_j)e^{i[\gamma_i(t)+\frac{\omega t}{2}-\gamma_j(t)-\frac{\omega t}{2}+ixv_i(t)-ixv_j(t)]}e^{-\frac{m\omega}{2\hbar}|x-x_i(t)|^2-\frac{m\omega}{2\hbar}|x-x_j(t)|^2}dx$$

Now, we reorganize the exponent of the exponential:

$$\begin{aligned}
&ix(v_i(t)-v_j(t)) - \frac{m\omega}{2\hbar}|x-x_i(t)|^2 - \frac{m\omega}{2\hbar}|x-x_j(t)|^2 = \\
&i\frac{m}{\hbar}x(\dot{x}_i(t)-\dot{x}_j(t)) - \frac{m\omega}{2\hbar}\left[2|x|^2 - 2x(x_i(t)+x_j(t)) + |x_i(t)|^2 + |x_j(t)|^2\right] = \\
&-\frac{m\omega}{2\hbar}\left[|x_i(t)|^2+|x_j(t)|^2\right] + \frac{m\omega}{4\hbar}\left((x_i(t)+x_j(t)-\frac{i}{\omega}(\dot{x}_i(t)-\dot{x}_j(t))\right)^2 \\
&-\frac{m\omega}{\hbar}\left[|x|^2 - x\left(x_i(t)+x_j(t)-\frac{i}{\omega}(\dot{x}_i(t)-\dot{x}_j(t))\right)\right] \\
&-\frac{m\omega}{\hbar}\left[\frac{1}{4}\left(x_i(t)+x_j(t)-\frac{i}{\omega}(\dot{x}_i(t)-\dot{x}_j(t)\right)^2\right] \\
&= -\frac{m\omega}{4\hbar}\left[2|x_i(t)|^2+2|x_j(t)|^2-|x_i(t)+x_j(t)|^2+\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{\omega^2}\right] \\
&-\frac{im}{2\hbar}\left(x_i(t)+x_j(t)\right)\left(\dot{x}_i(t)-\dot{x}_j(t)\right) \\
&-\frac{m\omega}{\hbar}\left[x-\frac{1}{2}\left(x_i(t)+x_j(t)-\frac{i}{\omega}(\dot{x}_i(t)-\dot{x}_j(t)\right)\right]^2 \\
&= -\frac{m\omega}{4\hbar}\left|x_i(t)-x_j(t)\right|^2 - \frac{m}{4\omega\hbar}|\dot{x}_i(t)-\dot{x}_j(t)|^2 \\
&\frac{m\omega}{\hbar}|y|^2 - \frac{im}{2\hbar}(x_i(t)+x_j(t))(\dot{x}_i(t)-\dot{x}_j(t)),
\end{aligned}$$

by changing variables to $y = x - \frac{x_i(t)+x_j(t)-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}$. So, we get

$$
\begin{aligned}
&Cov(\psi_i(t,x),\psi_j(t,x)) = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\Re\Big\{e^{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]}\times \\
&\times\int_{\mathbb{R}}\left(y+\frac{x_i(t)+x_j(t)-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}-x_i\right)\times \\
&\times\left(y+\frac{x_i(t)+x_j(t)-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}-x_j\right)e^{-\frac{m\omega}{\hbar}|y|^2}dy\Big\} = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\Re\Big\{e^{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]}\times \\
&\times\int_{\mathbb{R}}\left(y+\frac{-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}+\frac{x_i(t)-x_j(t)}{2}\right)\times \\
&\times\left(y+\frac{-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}-\frac{x_i(t)-x_j(t)}{2}\right)e^{-\frac{m\omega}{\hbar}|y|^2}dy\Big\} = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\Re\Big\{e^{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]}\times \\
&\times\int_{\mathbb{R}}\left(|y|^2-\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{4\omega^2}-\frac{1}{4}|x_i(t)-x_j(t)|^2-\frac{i}{\omega}y(\dot{x}_i(t)-\dot{x}_j(t))\right)e^{-\frac{m\omega}{\hbar}|y|^2}dy\Big\} = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\Re\Big\{e^{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]}\times \\
&\times\int_{\mathbb{R}}\left(|y|^2-\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{4\omega^2}-\frac{1}{4}|x_i(t)-x_j(t)|^2\right)e^{-\frac{m\omega}{\hbar}|y|^2}dy\Big\} = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\times \\
&\times\cos\Big\{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]\Big\}(J_1-J_2)
\end{aligned}
$$

with

$$
J_1=\int_{\mathbb{R}}y^2e^{-\frac{m\omega}{\hbar}y^2}dy.
$$

and

$$
J_2=-\left(\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{4\omega^2}+\frac{1}{4}|x_i(t)-x_j(t)|^2\right)\int_{\mathbb{R}}e^{-\frac{m\omega}{\hbar}y^2}dy.
$$

In both J_1 and J_2, we use the change of variables $z=y\left(\frac{m\omega}{\pi\hbar}\right)^{1/2}$, which implies $y=\left(\frac{m\omega}{\pi\hbar}\right)^{-1/2}z$ and so $dy=\left(\frac{m\omega}{\pi\hbar}\right)^{-1/2}dz$. This leads to

$$
J_1=\int_{\mathbb{R}}y^2e^{-\frac{m\omega}{\hbar}y^2}dy=\left(\frac{m\omega}{\hbar}\right)^{-1}\int_{\mathbb{R}}z^2e^{-z^2}\left(\frac{m\omega}{\hbar}\right)^{-1/2}dz=\left(\frac{m\omega}{\hbar}\right)^{-3/2}\frac{\pi^{1/2}}{2}
$$

and to

$$J_2 = \int_{\mathbb{R}} \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{4\omega^2} + \frac{1}{4}|x_i(t) - x_j(t)|^2 \right) e^{-\frac{m\omega}{\hbar}y^2} dy \tag{3.19}$$

$$= \left(\frac{m\omega}{\hbar}\right)^{-1/2} \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{4\omega^2} + \frac{1}{4}|x_i(t) - x_j(t)|^2 \right) \int_{\mathbb{R}} e^{-z^2} dz \tag{3.20}$$

$$= \left(\frac{m\omega}{\hbar}\right)^{-1/2} \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{4\omega^2} + \frac{1}{4}|x_i(t) - x_j(t)|^2 \right) \pi^{\frac{1}{2}}. \tag{3.21}$$

Now, putting everything together, we get

$$\begin{aligned}
&Cov(\psi_i(t,x), \psi_j(t,x)) = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \\
&\times \cos\left\{ i[\gamma_i(t) - \gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t) - \dot{x}_j(t))(x_i(t) + x_j(t))] \right\} (J_1 - J_2) = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{1/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \\
&\times \cos\left\{ i[\gamma_i(t) - \gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t) - \dot{x}_j(t))(x_i(t) + x_j(t))] \right\} \times \\
&\times \left(\left(\frac{m\omega}{\hbar}\right)^{-3/2} \frac{\pi^{1/2}}{2} - \left(\frac{m\omega}{\hbar}\right)^{-1/2} \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{4\omega^2} + \frac{1}{4}|x_i(t) - x_j(t)|^2 \right) \pi^{\frac{1}{2}} \right) = \\
&e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \\
&\times \frac{1}{2} \cos\left\{ [\gamma_i(t) - \gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t) - \dot{x}_j(t))(x_i(t) + x_j(t))] \right\} \times \\
&\times \left(\frac{\hbar}{m\omega} - \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{2\omega^2} + \frac{1}{2}|x_i(t) - x_j(t)|^2 \right) \right).
\end{aligned}$$

This completes the proof. □

Remark 3.2.25. *This* Measure of Amalgamation *seems to depend on time, but it actually does not as we will see in the proof of Theorem 3.2.1 (the non-stationary case), just after the next corollary. A crucial role is played by the explicit formula of the solutions of the* Classical Harmonic Oscillator *(3.5).*

As in the stationary case, we can deduce the following.

Corollary 3.2.26. *Consider two* moving solitons*:*

$$\psi_i(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_i(t) + \gamma_i(t) + \frac{m\omega t}{\hbar}]} e^{-\frac{m\omega}{2\hbar}|x - x_j(t)|^2},$$

and

$$\psi_j(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_j(t)+\gamma_j(t)+\frac{m\omega t}{2\hbar}]} e^{-\frac{m\omega}{2\hbar}|x-x_j(t)|^2},$$

for $1 \le i \le j \le N$ *and with* $\mathbf{x}(t)$, $\mathbf{v}(t)$ *and* $\gamma(t)$ *as in Theorem 3.2.8. Then, the* Measure of Correlation *between these two solitons is given by:*

$$\begin{aligned} Corr(\psi_i(t,x),\psi_j(t,x)) = \quad & e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \\ & \times \cos\left\{ i[\gamma_i(t)-\gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))] \right\} \times \\ & \times \left(1 - \frac{m\omega}{\hbar}\left(\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{2\omega^2} + \frac{1}{2}|x_i(t)-x_j(t)|^2\right)\right). \end{aligned}$$

Proof. It follows the same line of the stationary case and so we omit it. □

Now, we are ready to prove that, in the non-stationary case and for solitons, the *Simpson's Paradox* is persistent in time for the *Quantum Harmonic Oscillator*.

Proof of Theorem 3.2.1-Non-stationary case. The proof consists in showing that for every $t \in \mathbb{R}$

$$Cov(\psi_i(t,x),\psi_j(t,x)) = Cov(\psi_i(0,x),\psi_j(0,x))$$

and then conclude by the bi-linearity of the *measure of amalgamation*. We recall that

$$\begin{aligned} & Cov(\psi_i(t,x),\psi_j(t,x)) = \\ & e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \\ & \times \frac{1}{2}\cos\left\{ [\gamma_i(t)-\gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))] \right\} \times \\ & \times \left(\frac{\hbar}{m\omega} - \left(\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{2\omega^2} + \frac{1}{2}|x_i(t)-x_j(t)|^2\right)\right). \end{aligned}$$

At this point, we just need to compute each time dependent term and see that it is the same as the one at time $t = 0$. This is done taking advantage of the explicit form of the solutions of the *Harmonic Oscillator* (3.5). First of all, we know that $x(t)$ is of the form

$$x(t) = \alpha\cos(\omega t) + \frac{\beta}{\omega}\sin(\omega t)$$

for some $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ and so that

$$\dot{x}(t) = -\omega\alpha\sin(\omega t) + \beta\cos(\omega t).$$

Therefore,

$$
\begin{aligned}
&|x^i(t) - x^j(t)|^2 + \frac{1}{\omega^2}|\dot{x}^i(t) - \dot{x}^j(t)|^2 \\
&= |\alpha^i \cos(\omega t) + \frac{\beta^i}{\omega}\sin(\omega t) - \alpha^j \cos(\omega t) - \frac{\beta^j}{\omega}\sin(\omega t)|^2 + \\
&\frac{1}{\omega^2}| - \omega\alpha^i \sin(\omega t) + \beta^i \cos(\omega t) + \omega\alpha^j \sin(\omega t) - \beta^j \cos(\omega t)|^2 = \\
&|\alpha^i - \alpha^j|^2 \cos^2(\omega t) + \frac{1}{\omega^2}|\beta^i - \beta^j|^2 \sin^2(\omega t) \\
&+\frac{2}{\omega}(\alpha^i - \alpha^j)(\beta^i - \beta^j)\cos(\omega t)\sin(\omega t) + \\
&|\alpha^i - \alpha^j|^2 \sin^2(\omega t) + \frac{1}{\omega^2}|\beta^i - \beta^j|^2 \cos^2(\omega t) \\
&-\frac{2}{\omega}(\alpha^i - \alpha^j)(\beta^i - \beta^j)\cos(\omega t)\sin(\omega t) = \\
&|\alpha^i - \alpha^j|^2 + \frac{1}{\omega^2}|\beta^i - \beta^j|^2.
\end{aligned}
$$

Therefore, we have that

$$
\begin{aligned}
&e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \left(\frac{\hbar}{m\omega} - \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{2\omega^2} + \frac{1}{2}|x_i(t) - x_j(t)|^2\right)\right) = \\
&e^{-\frac{m\omega}{4\hbar}\left(|\alpha^i-\alpha^j|^2+\frac{1}{\omega^2}|\beta^i-\beta^j|^2\right)} \times \left(\frac{\hbar}{m\omega} - \frac{1}{2}\left(|\alpha^i - \alpha^j|^2 + \frac{1}{\omega^2}|\beta^i - \beta^j|^2\right)\right).
\end{aligned}
$$

Now, we have to concentrate on the phase:

$$
\cos\left\{[\gamma_i(t) - \gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t) - \dot{x}_j(t))(x_i(t) + x_j(t))]\right\}.
$$

First of all,

$$
\begin{aligned}
&(\dot{x}_i(t) - \dot{x}_j(t))(x_i(t) + x_j(t)) = \\
&\left[-\omega\alpha^i \sin(\omega t) + \beta^i \cos(\omega t) + \omega\alpha^j \sin(\omega t) - \beta^j \cos(\omega t)\right] \times \\
&\times \left[\alpha^i \cos(\omega t) + \frac{\beta^i}{\omega} \sin(\omega t) + \alpha^j \cos(\omega t) + \frac{\beta^j}{\omega} \sin(\omega t)\right] \\
&= \left[-\omega \sin(\omega t)(\alpha^i - \alpha^j) + \cos(\omega t)(\beta^i - \beta^j)\right] \\
&\times \left[(\alpha^i + \alpha^j) \cos(\omega t) + \frac{1}{\omega}(\beta^i + \beta^j) \sin(\omega t)\right] = \\
&\omega \sin(\omega t) \cos(\omega t) \left[-|\alpha^i|^2 + |\alpha^j|^2\right] + \frac{1}{\omega} \sin(\omega t) \cos(\omega t) \left[|\beta^i|^2 - |\beta^j|^2\right] + \\
&\cos^2(\omega t) \left[\alpha^i + \alpha^j\right] \left[\beta^i - \beta^j\right] + \sin^2(\omega t) \left[\alpha^j - \alpha^i\right] \left[\beta^i + \beta^j\right] = \\
&\frac{\sin(2\omega t)}{2} \left[\omega \left[|\alpha^j|^2 - |\alpha^i|^2\right] + \frac{1}{\omega} \left[|\beta^i|^2 - |\beta^j|^2\right]\right] + \\
&\alpha^j \beta^i - \beta^i \alpha^j + \left[\alpha^i \beta^i - \beta^j \alpha^j\right] \cos(2\omega t).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\frac{m}{2\hbar}(\dot{x}_i(t) - \dot{x}_j(t))(x_i(t) + x_j(t)) \\
&= \frac{m\omega}{4\hbar} \left[\left[|\alpha^j|^2 - |\alpha^i|^2\right] + \frac{1}{\omega^2} \left[|\beta^i|^2 - |\beta^j|^2\right]\right] \sin(2\omega t) + \\
&\frac{m}{2\hbar} \left[\alpha^j \beta^i - \beta^i \alpha^j\right] + \frac{m}{2\hbar} \left[\alpha^i \beta^i - \beta^j \alpha^j\right] \cos(2\omega t).
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\hbar\dot{\gamma}(t) = \frac{1}{2} m |\dot{x}(t)|^2 - \frac{1}{2} m\omega^2 |x(t)|^2 = \\
&\frac{1}{2} m \left[-\omega\alpha \sin(\omega t) + \beta \cos(\omega t)\right]^2 - \frac{1}{2} m\omega^2 \left[\alpha \cos(\omega t) + \frac{\beta}{\omega} \sin(\omega t)\right]^2 = \\
&\frac{1}{2} m \left[\omega^2 |\alpha|^2 \sin^2(\omega t) + |\beta|^2 \cos^2(\omega t) - 2\omega\alpha\beta \sin(\omega t) \cos(\omega t)\right] \\
&-\frac{1}{2} m\omega^2 \left[|\alpha|^2 \cos^2(\omega t) + \frac{|\beta|^2}{\omega^2} \sin^2(\omega t) + \frac{2}{\omega} \alpha\beta \sin(\omega t) \cos(\omega t)\right] = \\
&\frac{1}{2} m\omega^2 \left[-|\alpha|^2 \cos(2\omega t) + \frac{|\beta|^2}{\omega^2} \cos(2\omega t) - \frac{2}{\omega} \alpha\beta \sin(2\omega t)\right].
\end{aligned}
$$

Now, we integrate with respect to time side by side and obtain:

$$\hbar\left[\gamma(t)-\gamma(0)\right]=$$
$$\frac{1}{2}m\omega^2\left[-\frac{1}{2\omega}|\alpha|^2\sin(2\omega t)+\frac{1}{2\omega}\frac{|\beta|^2}{\omega^2}\sin(2\omega t)+\frac{1}{\omega^2}\alpha\beta\cos(2\omega t)\right]$$

and so

$$\gamma^i(t)=\gamma_0^i+\frac{m\omega}{4\hbar}\left[-|\alpha^i|^2+\frac{|\beta^i|^2}{\omega^2}\right]\sin(2\omega t)+\frac{m}{2\hbar}\alpha^i\beta^i\cos(2\omega t),$$

which implies that

$$\gamma^i(t)-\gamma^j(t)=\gamma_0^i-\gamma_0^j+\frac{m\omega}{4\hbar}\left[\left[|\alpha^j|^2-|\alpha^i|^2\right]+\frac{1}{\omega^2}\left[|\beta^i|^2-|\beta^j|^2\right]\right]\sin(2\omega t)$$
$$+\frac{m}{2\hbar}\left[\alpha^i\beta^i-\beta^j\alpha^j\right]\cos(2\omega t).$$

Therefore, putting everything together we get:

$$\gamma^i(t)-\gamma^j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))=\gamma_0^i-\gamma_0^j-\frac{m}{2\hbar}\left[\alpha^j\beta^i-\alpha^i\beta^j\right].$$

This is independent of time and so also

$$\cos\left(\gamma^i(t)-\gamma^j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))\right)$$
$$=\cos\left(\gamma_0^i-\gamma_0^j-\frac{m}{2\hbar}\left[\alpha^j\beta^i-\alpha^i\beta^j\right]\right).$$

is independent of time.

Remark 3.2.27. *It is interesting to notice that, since our* Measure of Amalgamation *basically measures the "angles" between the solitons, this measure ends up depending just on the initial angular momentum $\alpha^j\beta^i-\alpha^i\beta^j$, with $\alpha^i=x_0^i$ and $\beta^j=v_0^j$ and the initial complex phase $\gamma_0^i-\gamma_0^j$ for what concerns the angular part.*

Therefore, our *Measure of Amalgamation* is constant in time:

$$
\begin{aligned}
&Cov(\psi_i(t,x),\psi_j(t,x)) = e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\times\\
&\times\frac{1}{2}\cos\left\{[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]\right\}\times\\
&\times\left(\frac{\hbar}{m\omega}-\left(\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{2\omega^2}+\frac{1}{2}|x_i(t)-x_j(t)|^2\right)\right)=\\
&\frac{1}{2}\cos\left(\gamma_0^i-\gamma_0^j-\frac{m}{2\hbar}\left[\alpha^j\beta^i-\alpha^i\beta^j\right]\right)\times\\
&\times e^{-\frac{m\omega}{4\hbar}\left(|\alpha^i-\alpha^j|^2+\frac{1}{\omega^2}|\beta^i-\beta^j|^2\right)}\times\left(\frac{\hbar}{m\omega}-\frac{1}{2}\left(|\alpha^i-\alpha^j|^2+\frac{1}{\omega^2}|\beta^i-\beta^j|^2\right)\right)=\\
&Cov(\psi_i(0,x),\psi_j(0,x)).
\end{aligned}
$$

Finally, we have that

$$Cov(\psi_1(t,x),\psi_2(t,x)) = Cov(\psi_1(0,x),\psi_2(0,x)),$$

that

$$Cov(\psi_3(t,x),\psi_4(t,x)) = Cov(\psi_3(0,x),\psi_4(0,x))$$

and that

$$Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x)) = Cov(\psi_1(0,x)+\psi_3(0,x),\psi_2(0,x)+\psi_4(0,x)).$$

This implies that, if the *Simpson's paradox* appears at any time $t = t_0$, then it appears at any time $t \neq 0$, and that, if the *Simpson's paradox* does not appear at any time $t = t_0$, then it does not appear at any time $t \neq 0$. This completes the proof of Theorem 3.2.1 in the non-stationary case.

□

Remark 3.2.28. *A by-product of the following analysis is that the* Measure of Amalgamation *that we use is a conserved quantity, if we restrict our attention to solutions like solitons.*

3.2.3.4 The case $n \geq 2$

In spatial dimension $n \geq 2$, we can use a decomposition in *Spherical Harmonics* to obtain all possible solutions with *Schwartz regularity* and hence the *ground state*. This is a very well known result and report it here for completeness. We refer to [13], for an extended discussion on it.

Proposition 3.2.29. *The normalized* ground state *solution of equation (3.1)*

$$i\hbar\frac{\partial}{\partial t}\psi(t,x) = -\frac{\hbar^2}{2m}\Delta\psi(t,x) + \frac{1}{2}m\omega^2|x|^2\psi(t,x) \quad (t,x) \in (0,+\infty)\times\mathbb{R}^n.$$

in spatial dimension $n \geq 2$ is

$$\psi(t,\mathbf{x}) = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{n}{2}} e^{-\frac{m\omega}{2\hbar}|\mathbf{x}|^2} e^{i\frac{n\omega t}{2}}, \tag{3.22}$$

up to symmetries.

Proof. First of all, we have that in *spherical coordinates* the Laplacian can be rewritten in the following way:

$$\Delta_{\mathbf{x}} f = f'' + \frac{n-1}{r}f' + \frac{1}{r^2}\Delta_{\mathbf{S}^{n-1}} f,$$

where $r \in (0,+\infty)$, $f' := \frac{\partial}{\partial r}f$, $\Delta_{\mathbf{S}^{n-1}}$ is the *Laplace-Beltrami Operator* on $\mathbf{S}^{n-1}$. Since we are looking for the *ground state*, by the properties of the ground state developed in Section 3.2.2, we have

$$\psi(t,\mathbf{x}) = e^{it\lambda} f(r).$$

Plugging this inside (3.1), we get:

$$-\frac{\hbar^2}{2m}\left(f'' + \frac{n-1}{r}f'\right) + \frac{1}{2}m\omega^2|r|^2 f - \lambda\hbar f = 0.$$

Now, we plug inside the equation the *ansatz* $e^{-\frac{m\omega}{2\hbar}r^2}$ and one gets

$$-\frac{\hbar^2}{2m}\left(-\frac{m\omega}{\hbar} + r^2\left(\frac{m\omega}{\hbar}\right)^2 + \frac{n-1}{r}(-2r)\frac{m\omega}{2\hbar}\right) + \frac{1}{2}m\omega^2|r|^2 + \lambda\hbar = 0,$$

which implies $\lambda = \frac{n\omega}{2}$. And so one gets the solution

$$\psi(t,\mathbf{x}) = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{n}{4}} e^{-\frac{m\omega}{2\hbar}|\mathbf{x}|^2} e^{i\frac{n\omega t}{2}},$$

using a normalizing procedure completely similar to the case $n = 1$.

□

In a way similar to the case of spatial dimension $n = 1$, we can compute explicitly the value of the *Measure of Amalgamation*.

Proposition 3.2.30. *Consider two* moving solitons*:*

$$\psi_i(t,\mathbf{x}) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/4} e^{i[\mathbf{x}\cdot\mathbf{v}_i(t)+\gamma_i(t)+\frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|\mathbf{x}-\mathbf{x}_j(t)|^2},$$

and

$$\psi_j(t,\mathbf{x}) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/4} e^{i[\mathbf{x}\cdot\mathbf{v}_j(t)+\gamma_j(t)+\frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|\mathbf{x}-\mathbf{x}_j(t)|^2},$$

for $1 \leq i \leq j \leq N$ *and with* $\mathbf{x}(t)$*,* $\mathbf{v}(t)$ *and* $\gamma(t)$ *as in Theorem 3.2.8. Consider the case in which, for every* $t \in \mathbb{R}$*, one has that* $\mathbf{x}_k(t) = \mathbf{x}_k$*, for every* $k = 1, \ldots, N$ *independent of time. Then, the* Covariance *between these two solitons is given by:*

$$Cov(\psi_i(t,\mathbf{x}), \psi_j(t,\mathbf{x})) = \frac{1}{2}\cos(\gamma_i - \gamma_j)\left[\frac{n\hbar}{m\omega} - \frac{1}{2}|\mathbf{x}_i - \mathbf{x}_j|^2\right] e^{-\frac{m\omega}{4\hbar}|\mathbf{x}_i-\mathbf{x}_j|^2}. \qquad (3.23)$$

Proof. Again, we need to use the definition of Covariance and the shape of the solitons.

$$\begin{aligned}
&Cov(\psi_i(t,\mathbf{x}), \psi_j(t,\mathbf{x})) = \Re\int_{\mathbb{R}^n} (\mathbf{x}-\mu_i)(\mathbf{x}-\mu_j)^T \psi_i(t,\mathbf{x})\bar{\psi}_i(t,\mathbf{x})d\mathbf{x} \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} \Re\int_{\mathbb{R}^n} (\mathbf{x}-\mu_i)(\mathbf{x}-\mu_j) e^{i[\gamma_i+\frac{n\omega t}{2}-\gamma_j-\frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|\mathbf{x}-\mathbf{x}_i(t)|^2-\frac{m\omega}{2\hbar}|\mathbf{x}-\mathbf{x}_j(t)|^2} d\mathbf{x} \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} \Re e^{i[\gamma_i-\gamma_j]} \int_{\mathbb{R}^n} (\mathbf{x}-\mu_i)(\mathbf{x}-\mu_j) e^{-\frac{m\omega}{2\hbar}[2|\mathbf{x}|^2-2\mathbf{x}\cdot(\mathbf{x}_i+\mathbf{x}_j)+|\mathbf{x}_i|^2+|\mathbf{x}_j|^2]} d\mathbf{x} \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{2\hbar}[|\mathbf{x}_i|^2+|\mathbf{x}_j|^2]} \Re e^{i[\gamma_i-\gamma_j]} \times \\
&\times \int_{\mathbb{R}^n} (\mathbf{x}-\mu_i)(\mathbf{x}-\mu_j) e^{-\frac{m\omega}{\hbar}[|\mathbf{x}|^2-\mathbf{x}\cdot(\mathbf{x}_i+\mathbf{x}_j)+\frac{1}{4}|\mathbf{x}_i+\mathbf{x}_j|^2]+\frac{m\omega}{4\hbar}|\mathbf{x}_i+\mathbf{x}_j|^2} d\mathbf{x} \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{2\hbar}[|\mathbf{x}|_i^2+|\mathbf{x}_j|^2]+\frac{m\omega}{4\hbar}|\mathbf{x}_i+\mathbf{x}_j|^2} \cos(\gamma_i-\gamma_j) \\
&\times \int_{\mathbb{R}} (\mathbf{x}-\mu_i)(\mathbf{x}-\mu_j)^T e^{-\frac{m\omega}{\hbar}\left|\mathbf{x}-\frac{1}{2}(\mathbf{x}_i+\mathbf{x}_j)\right|^2} d\mathbf{x}.
\end{aligned}$$

Now, by changing variables to $\mathbf{y} = \mathbf{x} - \frac{\mathbf{x}_i-\mathbf{x}_j}{2}$, we get:

$$
\begin{aligned}
&Cov(\psi_i(t,\mathbf{x}),\psi_j(t,\mathbf{x})) = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}|\mathbf{x}_i-\mathbf{x}_j|^2}\cos(\gamma_i-\gamma_j) \\
&\times \int_{\mathbb{R}^n}\left(\mathbf{y}+\frac{\mathbf{x}_i-\mathbf{x}_j}{2}-\mathbf{x}_i\right)\left(\mathbf{y}+\frac{\mathbf{x}_i-\mathbf{x}_j}{2}-\mathbf{x}_j\right)e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2}d\mathbf{y} \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}|\mathbf{x}_i-\mathbf{x}_j|^2}\cos(\gamma_i-\gamma_j)\int_{\mathbb{R}^n}\left(|\mathbf{y}|^2-\frac{|\mathbf{x}_i-\mathbf{x}_j|^2}{4}\right)e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2}dy \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}|\mathbf{x}_i-\mathbf{x}_j|^2}\cos(\gamma_i-\gamma_j)\,(I_1-I_2)
\end{aligned}
$$

with

$$
I_1(n) = \int_{\mathbb{R}^n}|\mathbf{y}|^2 e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2}dy.
$$

and

$$
I_2(n) = \int_{\mathbb{R}^n}\frac{|\mathbf{x}_i-\mathbf{x}_j|^2}{4}e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2}d\mathbf{y}.
$$

In the computation of $I_1(n)$ and $I_2(n)$, there is a little bit of difference with respect to the case of spatial dimension $n=1$. In both $I_1(n)$ and $I_2(n)$, we use the change of variables $\mathbf{z}=\mathbf{y}\left(\frac{m\omega}{\hbar}\right)^{1/2}$, which implies $\mathbf{y}=\left(\frac{m\omega}{\hbar}\right)^{-1/2}\mathbf{z}$ and so $d\mathbf{y}=\left(\frac{m\omega}{\hbar}\right)^{-n/2}d\mathbf{z}$. This leads to

$$
\begin{aligned}
I_1(n) = \int_{\mathbb{R}^n}|\mathbf{y}|^2 e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2}d\mathbf{y} &= \left(\frac{m\omega}{\hbar}\right)^{-1}\int_{\mathbb{R}^n}|\mathbf{z}|^2 e^{-|\mathbf{z}|^2}\left(\frac{m\omega}{\hbar}\right)^{-n/2}d\mathbf{z} \\
&= \left(\frac{m\omega}{\hbar}\right)^{-1-\frac{n}{2}}\sum_{i=1}^{n} I^i(n),
\end{aligned}
$$

where

$$
I^i(n) := \int_{\mathbb{R}^n}|\mathbf{z}_i|^2 e^{-|\mathbf{z}|^2}d\mathbf{z} = \int_{\mathbb{R}}|\mathbf{z}_i|^2 e^{-|\mathbf{z}_i|^2}d\mathbf{z}_i \times \int_{\mathbb{R}^{n-1}} e^{-|\hat{\mathbf{z}}_i|^2}d\hat{\mathbf{z}}_i = \frac{\pi^{\frac{1}{2}}}{2}\pi^{\frac{n-1}{2}} = \frac{1}{2}\pi^{\frac{n}{2}}.
$$

Here $\hat{\mathbf{z}}_i$ is the vector $\mathbf{z}$ without the i-th component. Therefore,

$$
I_1(n) = \left(\frac{m\omega}{\pi\hbar}\right)^{-1-\frac{n}{2}}\Sigma_{i=1}^{n} I^i(n) = \left(\frac{m\omega}{\hbar}\right)^{-1-\frac{n}{2}}\frac{n}{2}\pi^{\frac{n}{2}} = \frac{n\hbar}{2m\omega}\left(\frac{\pi\hbar}{m\omega}\right)^{n/2}.
$$

With the same change of variables, we treat $I_2(n)$:

$$
\begin{aligned}
I_2(n) \quad &= \int_{\mathbb{R}^n} \frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{4} e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2} d\mathbf{y} \\
&= \left(\frac{m\omega}{\hbar}\right)^{-n/2} \frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{4} \int_{\mathbb{R}^n} e^{-|\mathbf{z}|^2} d\mathbf{z} = \left(\frac{m\omega}{\pi\hbar}\right)^{-n/2} \frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{4} \\
&= \frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{4} \left(\frac{\pi\hbar}{m\omega}\right)^{n/2}.
\end{aligned}
$$

Now, putting everything together, we get

$$
\begin{aligned}
&Cov(\psi_i(t,x), \psi_j(t,x)) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}|x_i - x_j|^2} \cos(\gamma_i - \gamma_j)\,(I_1 - I_2) \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}|x_i - x_j|^2} \cos(\gamma_i - \gamma_j) \left(\frac{n\hbar}{2m\omega}\left(\frac{\pi\hbar}{m\omega}\right)^{n/2} - \frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{4}\left(\frac{\pi\hbar}{m\omega}\right)^{n/2}\right) \\
&= \frac{1}{2}\cos(\gamma_i - \gamma_j)\left[\frac{n\hbar}{m\omega} - \frac{1}{2}|\mathbf{x}_i - \mathbf{x}_j|^2\right] e^{-\frac{m\omega}{4\hbar}|\mathbf{x}_i - \mathbf{x}_j|^2}.
\end{aligned}
$$

This completes the proof. □

Corollary 3.2.31. *Consider two* moving solitons*:*

$$
\psi_i(t,\mathbf{x}) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/4} e^{i[\mathbf{x}\cdot\mathbf{y}_i(t) + \gamma_i(t) + \frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|\mathbf{x} - \mathbf{x}_j(t)|^2},
$$

and

$$
\psi_j(t,\mathbf{x}) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/4} e^{i[\mathbf{x}\cdot\mathbf{v}_j(t) + \gamma_j(t) + \frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|\mathbf{x} - \mathbf{x}_j(t)|^2},
$$

for $1 \le i \le j \le N$ and with $\mathbf{x}(t)$, $\mathbf{v}(t)$ and $\gamma(t)$ as in Theorem 3.2.8. Consider the case in which, for every $t \in \mathbb{R}$, one has that $\mathbf{x}_k(t) = \mathbf{x}_k$, for every $k = 1, \ldots, N$ independent of time. Then, the Measure of Correlation *between these two solitons is given by:*

$$
Corr(\psi_i(t,\mathbf{x}), \psi_j(t,\mathbf{x})) = \cos(\gamma_i - \gamma_j)\left[1 - \frac{m\omega}{2n\hbar}|x_i - x_j|^2\right] e^{-\frac{m\omega}{4\hbar}|\mathbf{x}_i - \mathbf{x}_j|^2}.
$$

Proof. Since

$$
Cov(\psi_i(t,\mathbf{x}), \psi_j(t,\mathbf{x})) = \frac{1}{2}\cos(\gamma_i - \gamma_j)\left[\frac{n\hbar}{m\omega} - \frac{1}{2}|\mathbf{x}_i - \mathbf{x}_j|^2\right] e^{-\frac{m\omega}{4\hbar}|\mathbf{x}_i - \mathbf{x}_j|^2}
$$

and $Var(\psi_i(t,\mathbf{x})) = Var(\psi_j(t,\mathbf{x})) = \frac{n\hbar}{2m\omega}$, then, by definition of *Measure of Correlation* one gets

$$Corr(\psi_i(t,\mathbf{x}),\psi_j(t,\mathbf{x})) = \frac{Cov(\psi_i(t,\mathbf{x}),\psi_j(t,\mathbf{x}))}{Var(\psi_i(t,\mathbf{x}))^{\frac{1}{2}}Var(\psi_j(t,\mathbf{x}))^{\frac{1}{2}}}$$
$$= \cos(\gamma_i-\gamma_j)\left[1-\frac{m\omega}{2n\hbar}|\mathbf{x}_i-\mathbf{x}_j|^2\right]e^{-\frac{m\omega}{4\hbar}|\mathbf{x}_i-\mathbf{x}_j|^2}.$$

This completes the proof of the corollary. □

Now, we are ready to show that, according to the choice of the parameters, the *Simpson's Paradox* can occur also in spatial dimension $n > 1$.

Proof of Theorem 3.2.1. Again, the proof consists in finding parameters such that the *Simpson's Reversal* occurs, namely such that

$$Cov(\psi_1(t,x),\psi_2(t,x)) > 0,$$

$$Cov(\psi_3(t,x),\psi_4(t,x)) > 0$$

but

$$Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x)) < 0$$

or vice versa,

$$Cov(\psi_1(t,x),\psi_3(t,x)) < 0,$$

$$Cov(\psi_2(t,x),\psi_4(t,x)) < 0$$

but

$$Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x)) > 0.$$

We define

$$L_{ij}^2 := \frac{m\omega}{2n\hbar}|x_i-x_j|^2$$

and so $Cov(\psi_i(t,x),\psi_j(t,x))$ can be rewritten in the following way:

$$Cov(\psi_i(t,x),\psi_j(t,x)) = \frac{n\hbar}{2m\omega}\cos(\gamma_i-\gamma_j)\left[1-L_{ij}^2\right]e^{-\frac{n}{2}L_{ij}^2}.$$

This definition of L_{ij} is slightly different from the one given in the proof of the case $n = 1$. From now on, since we are just looking for the existence of a set of parameters which leads to the *Simpson's Paradox* we assume that x_1, x_2, x_3 and x_4 stay on a line. With this choice, the proof reduces to the case of spatial dimension $n = 1$. It is enough then to choose $L_{12} = 1-\epsilon_1^2$, $L_{34} = 1-\epsilon_2^2$, $L_{23} = 1+\delta^2$ with $\epsilon_1 << 1$, $\epsilon_2 << 1$, with $\epsilon_1 = \epsilon_2 = \alpha > 0$

and $\delta = k\alpha > 0$ for $\alpha^2 < \frac{2\log\left(\frac{k^2}{2}\right)}{n(k^2+1)}$ and $L_{14} = L_{12} + L_{23} + L_{34}$. It is enough to choose for example $\alpha = 10^{-100}$ and $k = 100$ and both $\alpha^2 < \frac{2\log\left(\frac{k^2}{2}\right)}{k^2+1}$ and $0 < k\alpha < 1$ are satisfied. This concludes the proof of the case $\gamma_i = \gamma_j$, for every $i, j = 1, \ldots, 4$. The proof in the case $\gamma_i \neq \gamma_j$, for every $i, j = 1, \ldots, 4$, $i \neq j$ is very similar, so we omit it. This completes the proof in the stationary case for spatial dimension $n > 1$.

□

Remark 3.2.32. *In the case $n = 1$, there is more constraint in the choice of the parameters, since, for example, x_1, x_2, x_3 and x_4 must be on the line and so L_{ij} must satisfy something like $L_{12} + L_{23} + L_{34} = L_{41}$. This is not anymore true in dimension $n > 1$, because the x_i's can stay on different geometrical objects like squares or triangles and still give rise to the* Simpson's Paradox*. This does not necessarily mean that, in higher dimension, the* Simpson's Paradox *is more likely to happen, because also the configurations of not occurrence of the* Simpson's Paradox *increase in number. See Section 3.2.8 and [94].*

Remark 3.2.33. *Also in the case of dimension $n > 1$, we can talk about* Semiclassical Regime*,* Uncorrelation Regime *and* Anti-Semiclassical Regime *but, this time, $L_{ij} > 1$ corresponds to the case $\hbar < \frac{m\omega}{2n}|x_i - x_j|^2$, $L_{ij} = 1$ to the case $\hbar = \frac{m\omega}{2n}|x_i - x_j|^2$ and $L_{ij} < 1$ to the case $\hbar > \frac{m\omega}{2n}|x_i - x_j|^2$. This means that, with the increase of the dimension n, the* Anti-Semiclassical Regime *is, in some sense, more likely to appear. See again Section 3.2.8 for more details on some probabilistic issues related to the* Simpson's Paradox*.*

3.2.3.5 The non-stationary case $n \geq 2$

Now, we pass to the non-stationary case in spatial dimension $n \geq 2$. Consider moving solitons of the form

$$\psi_i(t, \mathbf{x}) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/4} e^{i[\mathbf{x}\cdot v_i(t) + \gamma_i(t) + \frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|\mathbf{x}|^2}, \tag{3.24}$$

for $i = 1, \ldots, N$ with N the number of solitons and with $\mathbf{x}(t)$, $\mathbf{v}(t)$ and $\gamma(t)$ as in Theorem 3.2.8. Again, for these objects, we compute how the *center of Mass* moves and how the *Variance* changes under the evolution of (3.1).

Lemma 3.2.34. *The* center of Mass *of each $\psi_i(t, \mathbf{x})$ as defined in (3.13) is given by*

$$\mu_i := \int_{\mathbb{R}} \mathbf{x}|\psi(t, \mathbf{x})|^2 dx = \mathbf{x}_i(t), \tag{3.25}$$

for every $i = 1, \ldots, N$.

Proof. We first recall the following fact about integrals involving *Gaussians*:

$$\int_{\mathbb{R}^n} e^{-a|\mathbf{x}|^2} d\mathbf{x} = \left(\int_{\mathbb{R}} e^{-ax^2} dx\right)^n = \left(\frac{\pi}{a}\right)^{n/2},$$

for every $a > 0$. Therefore, for every $i = 1, \ldots, N$ and using the change of variables $\mathbf{y} := \mathbf{x} - \mathbf{x}(t)$, we have

$$\begin{aligned}
\mu_i &:= \int_{\mathbb{R}} \mathbf{x}|\psi(t, \mathbf{x})|^2 d\mathbf{x} = \left(\frac{m\omega}{\pi\hbar}\right)^{1/2} \int_{\mathbb{R}^n} \mathbf{x} e^{-\frac{m\omega}{\hbar}[\mathbf{x}-\mathbf{x}_i(t)]^2} d\mathbf{x} \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} \int_{\mathbb{R}^n} [\mathbf{x} + \mathbf{x}_i(t)] e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2} d\mathbf{y} \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} \int_{\mathbb{R}^n} \mathbf{x}(t) e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2} d\mathbf{y} = \mathbf{x}_i(t).
\end{aligned}$$

□

Lemma 3.2.35. *The* Variance *of each $\psi_i(t, \mathbf{x})$ as defined in (3.13) is given by*

$$Var(\psi_i(t, \mathbf{x})) := \int_{\mathbb{R}^n} |\mathbf{x}|^2 |\psi(t, \mathbf{x})|^2 d\mathbf{x} - \left(\int_{\mathbb{R}^n} \mathbf{x}|\psi(t, \mathbf{x})|^2 dx\right)^2 = \frac{n\hbar}{2m\omega}, \tag{3.26}$$

for every $i = 1, \ldots, N$.

Proof. First of all, we notice that

$$\left(\int_{\mathbb{R}^n} \mathbf{x}|\psi(t, \mathbf{x})|^2 d\mathbf{x}\right)^2 = |\mu_i|^2 = |\mathbf{x}_i(t)|^2$$

and so we just need to compute $\int_{\mathbb{R}} |x|^2 |\psi(t, x)|^2 dx$.

$$
\begin{aligned}
Var(\psi_i(t,\mathbf{x})) &:= -|\mathbf{x}_i(t)|^2 + \int_{\mathbb{R}^n} |\mathbf{x}|^2 |\psi(t,\mathbf{x})|^2 d\mathbf{x} \\
&= -|\mathbf{x}_i(t)|^2 + \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} \int_{\mathbb{R}^n} |\mathbf{x}|^2 e^{-\frac{m\omega}{\hbar}[\mathbf{x}-\mathbf{x}_i(t)]^2} d\mathbf{x} \\
&= -|\mathbf{x}_i(t)|^2 + \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} \int_{\mathbb{R}^n} [|\mathbf{y}|^2 + 2\mathbf{x}_i(t)\cdot\mathbf{y} + |\mathbf{x}_i(t)|^2] e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2} d\mathbf{y} \\
&= -|\mathbf{x}_i(t)|^2 + \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} \int_{\mathbb{R}^n} [|\mathbf{y}|^2 + |\mathbf{x}_i(t)|^2] e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2} d\mathbf{y} \\
&= (1-1)\left(\frac{m\omega}{\pi\hbar}\right)^{n/2} |\mathbf{x}_i(t)|^2 + \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} \int_{\mathbb{R}^n} |\mathbf{y}|^2 e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2} d\mathbf{y} \\
&= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2} \int_{\mathbb{R}^n} |\mathbf{y}|^2 e^{-\frac{m\omega}{\hbar}|\mathbf{y}|^2} d\mathbf{y} \\
&= \left(\frac{m\omega}{\hbar}\right)^{n/2-1-n/2} \pi^{-n/2} \int_{\mathbb{R}^n} |\mathbf{z}|^2 e^{-|\mathbf{z}|^2} d\mathbf{z} = \frac{n\hbar}{2m\omega}.
\end{aligned}
$$

□

Now, we can compute how the *Measure of Correlation* between two solitons vary over time also in the case $n > 1$.

Proposition 3.2.36. *Consider two* moving solitons*:*

$$\psi_i(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/4} e^{i[x\cdot v_i(t)+\gamma_i(t)+\frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x-x_j(t)|^2},$$

and

$$\psi_j(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/4} e^{i[x\cdot v_j(t)+\gamma_j(t)+\frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x-x_j(t)|^2},$$

for $1 \le i \le j \le N$ *and with* $\mathbf{x}(t)$, $\mathbf{v}(t)$ *and* $\gamma(t)$ *as in Theorem 3.2.8. Then, the* Covariance *between these two solitons is given by:*

$$Cov(\psi_i(t,x),\psi_j(t,x)) = e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \tag{3.27}$$

$$\times \frac{1}{2}\cos\left\{i[\gamma_i(t)-\gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]\right\} \times \tag{3.28}$$

$$\times \left(\frac{n\hbar}{m\omega} - \left(\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{2\omega^2} + \frac{1}{2}|x_i(t)-x_j(t)|^2\right)\right). \tag{3.29}$$

Proof. Again, we need to use the definition of Covariance and the shape of the solitons.

$$Cov(\psi_i(t,x),\psi_j(t,x)) = \Re\int_{\mathbb{R}^n}(\mathbf{x}-\mu_i)(\mathbf{x}-\mu_j)\psi_i(t,\mathbf{x})\overline{\psi}_i(t,\mathbf{x})d\mathbf{x}$$
$$= \left(\frac{m\omega}{\pi\hbar}\right)^{n/2}\times$$
$$\times\Re\int_{\mathbb{R}}(x-\mu_i)(x-\mu_j)e^{i[\gamma_i(t)+\frac{n\omega t}{2}-\gamma_j(t)-\frac{n\omega t}{2}+ixv_i(t)-ixv_j(t)]}e^{-\frac{m\omega}{2\hbar}|x-x_i(t)|^2-\frac{m\omega}{2\hbar}|x-x_j(t)|^2}dx$$

Now, we reorganize the exponent of the exponential. Similarly to the case $n=1$:

$$ix(v_i(t)-v_j(t))-\frac{m\omega}{2\hbar}|x-x_i(t)|^2-\frac{m\omega}{2\hbar}|x-x_j(t)|^2 =$$
$$-\frac{m\omega}{4\hbar}\,|x_i(t)-x_j(t)|^2-\frac{m}{4\omega\hbar}|\dot{x}_i(t)-\dot{x}_j(t)|^2-\frac{m\omega}{\hbar}|y|^2$$
$$-\frac{im}{2\hbar}(x_i(t)+x_j(t))(\dot{x}_i(t)-\dot{x}_j(t)),$$

by changing variables to $y = x-\frac{x_i(t)+x_j(t)-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}$. So, we get

$$
\begin{aligned}
&Cov(\psi_i(t,x),\psi_j(t,x)) = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\Re\Big\{e^{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]}\times \\
&\times\int_{\mathbb{R}}\left(y+\frac{x_i(t)+x_j(t)-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}-x_i\right)\times \\
&\times\left(y+\frac{x_i(t)+x_j(t)-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}-x_j\right)e^{-\frac{m\omega}{\hbar}|y|^2}dy\Big\} = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\Re\Big\{e^{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]}\times \\
&\times\int_{\mathbb{R}}\left(y+\frac{-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}+\frac{x_i(t)-x_j(t)}{2}\right)\times \\
&\times\left(y+\frac{-\frac{i}{\omega}[\dot{x}_i(t)-\dot{x}_j(t)]}{2}-\frac{x_i(t)-x_j(t)}{2}\right)e^{-\frac{m\omega}{\hbar}|y|^2}dy\Big\} = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\Re\Big\{e^{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]}\times \\
&\times\int_{\mathbb{R}}\left(|y|^2-\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{4\omega^2}-\frac{1}{4}|x_i(t)-x_j(t)|^2-\frac{i}{\omega}y(\dot{x}_i(t)-\dot{x}_j(t))\right)e^{-\frac{m\omega}{\hbar}|y|^2}dy\Big\} = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\Re\Big\{e^{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]}\times \\
&\times\int_{\mathbb{R}}\left(|y|^2-\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{4\omega^2}-\frac{1}{4}|x_i(t)-x_j(t)|^2\right)e^{-\frac{m\omega}{\hbar}|y|^2}dy\Big\} = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\times \\
&\times\cos\Big\{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]\Big\}(J_1-J_2)
\end{aligned}
$$

with

$$J_1=\int_{\mathbb{R}}y^2e^{-\frac{m\omega}{\hbar}y^2}dy.$$

and

$$J_2=-\left(\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{4\omega^2}+\frac{1}{4}|x_i(t)-x_j(t)|^2\right)\int_{\mathbb{R}}e^{-\frac{m\omega}{\hbar}y^2}dy.$$

In both J_1 and J_2, we use the change of variables $z=y\left(\frac{m\omega}{\pi\hbar}\right)^{1/2}$, which implies $y=\left(\frac{m\omega}{\pi\hbar}\right)^{-1/2}z$ and so $dy=\left(\frac{m\omega}{\pi\hbar}\right)^{-n/2}dz$. This leads to

$$J_1=\int_{\mathbb{R}}|y|^2e^{-\frac{m\omega}{\hbar}y^2}dy=\left(\frac{m\omega}{\hbar}\right)^{-1}\int_{\mathbb{R}}|z|^2e^{-z^2}\left(\frac{m\omega}{\hbar}\right)^{-n/2}dz=\left(\frac{m\omega}{\hbar}\right)^{-1-n/2}\frac{n}{2}\pi^{n/2}$$

and to

$$J_2 = \int_{\mathbb{R}} \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{4\omega^2} + \frac{1}{4}|x_i(t) - x_j(t)|^2 \right) e^{-\frac{m\omega}{\hbar}y^2} dy \tag{3.30}$$

$$= \left(\frac{m\omega}{\hbar}\right)^{-n/2} \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{4\omega^2} + \frac{1}{4}|x_i(t) - x_j(t)|^2 \right) \int_{\mathbb{R}} e^{-z^2} dz \tag{3.31}$$

$$= \left(\frac{m\omega}{\hbar}\right)^{-n/2} \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{4\omega^2} + \frac{1}{4}|x_i(t) - x_j(t)|^2 \right) \pi^{\frac{n}{2}}. \tag{3.32}$$

Now, putting everything together, we get

$$
\begin{aligned}
&Cov(\psi_i(t,x), \psi_j(t,x)) = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \\
&\times \cos\left\{ i[\gamma_i(t) - \gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t) - \dot{x}_j(t))(x_i(t) + x_j(t))] \right\} (J_1 - J_2) = \\
&\left(\frac{m\omega}{\pi\hbar}\right)^{n/2} e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \\
&\times \cos\left\{ i[\gamma_i(t) - \gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t) - \dot{x}_j(t))(x_i(t) + x_j(t))] \right\} \times \\
&\times \left(\left(\frac{m\omega}{\hbar}\right)^{-1-n/2} \frac{n}{2}\pi^{n/2} - \left(\frac{m\omega}{\hbar}\right)^{-n/2} \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{4\omega^2} + \frac{1}{4}|x_i(t) - x_j(t)|^2 \right) \pi^{\frac{1}{2}} \right) = \\
&e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2 - \frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2} \times \\
&\times \frac{1}{2} \cos\left\{ [\gamma_i(t) - \gamma_j(t) - \frac{m}{2\hbar}(\dot{x}_i(t) - \dot{x}_j(t))(x_i(t) + x_j(t))] \right\} \times \\
&\times \left(\frac{n\hbar}{m\omega} - \left(\frac{|\dot{x}_i(t) - \dot{x}_j(t)|^2}{2\omega^2} + \frac{1}{2}|x_i(t) - x_j(t)|^2 \right) \right).
\end{aligned}
$$

This completes the proof. □

As in the stationary case, we can deduce the following.

Corollary 3.2.37. *Consider two* moving solitons*:*

$$\psi_i(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/4} e^{i[x \cdot v_i(t) + \gamma_i(t) + \frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x - x_j(t)|^2},$$

and

$$\psi_j(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{n/4} e^{i[x \cdot v_j(t) + \gamma_j(t) + \frac{n\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x - x_j(t)|^2},$$

for $1 \leq i \leq j \leq N$ *and with* $\mathbf{x}(t)$*,* $\mathbf{v}(t)$ *and* $\gamma(t)$ *as in Theorem 3.2.8. Then, the* Measure of Correlation *between these two solitons is given by:*

$$Corr(\psi_i(t,x),\psi_j(t,x)) = e^{-\frac{m\omega}{4\hbar}[x_i(t)-x_j(t)]^2-\frac{m}{4\omega\hbar}[\dot{x}_i(t)-\dot{x}_j(t)]^2}\times$$
$$\times\cos\left\{i[\gamma_i(t)-\gamma_j(t)-\frac{m}{2\hbar}(\dot{x}_i(t)-\dot{x}_j(t))(x_i(t)+x_j(t))]\right\}\times$$
$$\times\left(1-\frac{m\omega}{n\hbar}\left(\frac{|\dot{x}_i(t)-\dot{x}_j(t)|^2}{2\omega^2}+\frac{1}{2}|x_i(t)-x_j(t)|^2\right)\right).$$

Proof. It follows the same line of the stationary case and so we omit it. □

Proof of Theorem 3.2.1. The proof of the main theorem in the non-stationary case for spatial dimension $n>1$ follows the same strategy employed in the case $n=1$, since the quantities depending on time of $Cov(\psi_i(t,x),\psi_j(t,x))$ are the same and what varies are just some constants which now depend on the dimension n. Therefore, we have the persistence of the *Simpson's Paradox* also in the case $n>1$. This completes the proof of Theorem 3.2.1. □

3.2.4 Proof of Theorem 3.2.5

In this subsection, we give the complete proof of Theorem 3.2.5. Now, we suppose that there exist solitons $\psi_i(t,x)$ for $i=1,\dots,4$ such that

$$Cov(\psi_1(t,x),\psi_2(t,x))>0,$$

$$Cov(\psi_3(t,x),\psi_4(t,x))>0$$

but

$$Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x))<0.$$

We want to prove that the same inequalities work if we consider $\tilde{\psi}_i(t,x)$ for $i=1,\dots,4$ with parameters $(\tilde{x}_i(t),\tilde{\gamma}_i(t),\tilde{v}_i(t))$ for $i=1,\dots,4$, where $\tilde{x}_i(t)=x_i(t)+\delta_i=x_i+\delta_i$, $\tilde{\gamma}_i(t)=\gamma_i(t)+\epsilon_i=\gamma_i+\epsilon_i$ and $\tilde{v}_i(t)=v_i(t)=0$ for every $i=1,\dots,4$. It will work in a similar way for the vice versa, namely when

$$Cov(\psi_1(t,x),\psi_3(t,x))<0,$$

$$Cov(\psi_2(t,x),\psi_4(t,x))<0$$

but

$$Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x))>0$$

and so we will omit the proof in this case.
We already know that

$$Cov(\psi_i(t,x),\psi_j(t,x)) = \frac{1}{2}\cos(\gamma_i-\gamma_j)\left[\frac{n\hbar}{m\omega}-\frac{1}{2}|x_i-x_j|^2\right]e^{-\frac{m\omega}{4\hbar}|x_i-x_j|^2}$$

and also that

$$Cov(\tilde{\psi}_i(t,x),\tilde{\psi}_j(t,x)) = \frac{1}{2}\cos(\tilde{\gamma}_i-\tilde{\gamma}_j)\left[\frac{n\hbar}{m\omega}-\frac{1}{2}|\tilde{x}_i-\tilde{x}_j|^2\right]e^{-\frac{m\omega}{4\hbar}|\tilde{x}_i-\tilde{x}_j|^2}.$$

Now, we expand this last expression and get

$$\begin{aligned}
&Cov(\tilde{\psi}_i(t,x),\tilde{\psi}_j(t,x)) = \frac{1}{2}\cos(\tilde{\gamma}_i-\tilde{\gamma}_j)\left[\frac{n\hbar}{m\omega}-\frac{1}{2}|\tilde{x}_i-\tilde{x}_j|^2\right]e^{-\frac{m\omega}{4\hbar}|\tilde{x}_i-\tilde{x}_j|^2} = \\
&= \frac{1}{2}\cos(\gamma_i+\epsilon_i-\gamma_j-\epsilon_j)\left[\frac{n\hbar}{m\omega}-\frac{1}{2}|x_i+\delta_i-x_j-\delta_j|^2\right]e^{-\frac{m\omega}{4\hbar}|x_i+\delta_i-x_j-\delta_j|^2} = \\
&\frac{1}{2}\left[\cos(\gamma_i-\gamma_j)\cos(+\epsilon_i-\epsilon_j)-\sin(\gamma_i-\gamma_j)\sin(+\epsilon_i-\epsilon_j)\right]\times \\
&\times\left[\frac{n\hbar}{m\omega}-\frac{1}{2}|x_i-x_j|^2-(\delta_i-\delta_j)\cdot(x_i-x_j)+\frac{1}{2}|\delta_i-\delta_j|^2\right]\times \\
&\times e^{-\frac{m\omega}{4\hbar}\left[|x_i-x_j|^2+2(\delta_i-\delta_j)\cdot(x_i-x_j)+|\delta_i-\delta_j|^2\right]} = \\
&\cos(\epsilon_i-\epsilon_j)Cov(\psi_i(t,x),\psi_j(t,x))e^{-\frac{m\omega}{4\hbar}\left[2(\delta_i-\delta_j)\cdot(x_i-x_j)+|\delta_i-\delta_j|^2\right]} + \\
&\frac{1}{2}\left[-\sin(\gamma_i-\gamma_j)\sin(+\epsilon_i-\epsilon_j)\right]\times \\
&\times\left[\frac{n\hbar}{m\omega}-\frac{1}{2}|x_i-x_j|^2-(\delta_i-\delta_j)\cdot(x_i-x_j)+\frac{1}{2}|\delta_i-\delta_j|^2\right]\times \\
&\times e^{-\frac{m\omega}{4\hbar}\left[|x_i-x_j|^2+2(\delta_i-\delta_j)\cdot(x_i-x_j)+|\delta_i-\delta_j|^2\right]} + \\
&\frac{1}{2}\left[\cos(\gamma_i-\gamma_j)\cos(+\epsilon_i-\epsilon_j)-\sin(\gamma_i-\gamma_j)\sin(+\epsilon_i-\epsilon_j)\right]\times \\
&\times\left[-(\delta_i-\delta_j)\cdot(x_i-x_j)+\frac{1}{2}|\delta_i-\delta_j|^2\right]\times \\
&\times e^{-\frac{m\omega}{4\hbar}\left[|x_i-x_j|^2+2(\delta_i-\delta_j)\cdot(x_i-x_j)+|\delta_i-\delta_j|^2\right]}
\end{aligned}$$

Now when $|\delta_i-\delta_j|\to 0$ and $|\epsilon_i-\epsilon_j|\to 0$, we have that the first term of the sum

$$\cos(\epsilon_i-\epsilon_j)Cov(\psi_i(t,x),\psi_j(t,x))e^{-\frac{m\omega}{4\hbar}\left[2(\delta_i-\delta_j)\cdot(x_i-x_j)+|\delta_i-\delta_j|^2\right]}\to Cov(\psi_i(t,x),\psi_j(t,x)),$$

while the second term of the sum

$$\frac{1}{2}\left[-\sin(\gamma_i-\gamma_j)\sin(+\epsilon_i-\epsilon_j)\right]\left[\frac{n\hbar}{m\omega}-\frac{1}{2}|x_i-x_j|^2-(\delta_i-\delta_j)(x_i-x_j)+\frac{1}{2}|\delta_i-\delta_j|^2\right]\to 0$$

and the third term of the sum

$$\frac{1}{2}\left[\cos(\gamma_i-\gamma_j)\cos(+\epsilon_i-\epsilon_j)-\sin(\gamma_i-\gamma_j)\sin(+\epsilon_i-\epsilon_j)\right]$$
$$\times\left[\frac{1}{2}|\delta_i-\delta_j|^2-(\delta_i-\delta_j)(x_i-x_j)\right]e^{-\frac{m\omega}{4\hbar}\left[|x_i-x_j|^2+2(\delta_i-\delta_j)\cdot(x_i-x_j)+|\delta_i-\delta_j|^2\right]}\to 0.$$

Therefore

$$Cov(\tilde{\psi}_i(t,x),\tilde{\psi}_j(t,x))\to Cov(\psi_i(t,x),\psi_j(t,x)).$$

Hence, when $|\delta_i-\delta_j|\le K_{ij}<<1$ and $|\epsilon_i-\epsilon_j|\le M_{ij}<<1$, if $Cov(\psi_i(t,x),\psi_j(t,x))\ge 0$, then $Cov(\tilde{\psi}_i(t,x),\tilde{\psi}_j(t,x))\ge 0$. Now

$$\begin{aligned}
&Cov(\tilde{\psi}_1(t,x)+\tilde{\psi}_3(t,x),\tilde{\psi}_2(t,x)+\tilde{\psi}_4(t,x))=\\
&Cov(\tilde{\psi}_1(t,x),\tilde{\psi}_2(t,x))+Cov(\tilde{\psi}_1(t,x),\tilde{\psi}_4(t,x))+\\
&Cov(\tilde{\psi}_3(t,x),\tilde{\psi}_2(t,x))+Cov(\tilde{\psi}_3(t,x),\tilde{\psi}_4(t,x))\\
&\to\\
&Cov(\psi_1(t,x),\psi_2(t,x))+Cov(\psi_1(t,x),\psi_4(t,x))+\\
&Cov(\psi_3(t,x),\psi_2(t,x))+Cov(\psi_3(t,x),\psi_4(t,x))\\
&=Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x))<0.
\end{aligned}$$

Therefore, if we choose

$$K<\frac{1}{4}\min(K_{12},K_{14},K_{32},K_{34})$$

and

$$M<\frac{1}{4}\min(M_{12},M_{14},M_{32},M_{34}),$$

every soliton with parameters in the ball

$$B_{KM}:=\{(\tilde{x}_i,\tilde{\gamma}_i):\sup_{i=1,\dots,4}|\tilde{x}_i-x_i|\le K,\ \sup_{i=1,\dots,4}|\tilde{\gamma}_i-\gamma_i|\le M\}$$

present the *Simpson's Paradox*. This proves the first part of the theorem.
Now, we pass to the second part of the theorem. Consider solutions of equation (3.1) of the form $\tilde{\psi}_i(t,x)=\psi_i(t,x)+w_i(t,x)$, where $\psi(t,x)=\Sigma_{i=1}^4\psi_i(t,x)$ is a soliton solution

such that, at time $\tilde{t}$, *Simpson's Paradox* occurs. Now, by explicit computation, we get:

$$
\begin{aligned}
&Cov(\tilde{\psi}_i(\tilde{t},x),\tilde{\psi}_j(\tilde{t},x)) = Cov(\psi_i(\tilde{t},x)+w_i(\tilde{t},x),\psi_j(\tilde{t},x)+w_j(\tilde{t},x)) = \\
&Cov(\psi_i(\tilde{t},x),\psi_j(\tilde{t},x)) + Cov(w_i(\tilde{t},x),\psi_j(\tilde{t},x)) + \\
&+Cov(\psi_i(\tilde{t},x),w_j(\tilde{t},x)) + Cov(w_i(\tilde{t},x),w_j(\tilde{t},x)).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\Big|Cov(\tilde{\psi}_i(\tilde{t},x),\tilde{\psi}_j(\tilde{t},x)) - Cov(\psi_i(\tilde{t},x),\psi_j(\tilde{t},x))\Big| \\
&= \big|Cov(w_i(\tilde{t},x),\psi_j(\tilde{t},x)) + Cov(\psi_i(\tilde{t},x),w_j(\tilde{t},x)) + Cov(w_i(\tilde{t},x),w_j(\tilde{t},x))\big| \leq \\
&\leq \big|Cov(w_i(\tilde{t},x),\psi_j(\tilde{t},x))\big| + \big|Cov(\psi_i(\tilde{t},x),w_j(\tilde{t},x))\big| + \big|Cov(w_i(\tilde{t},x),w_j(\tilde{t},x))\big| \\
&\leq \Big(\big|\big|x\psi_j(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)} + \big|\big|\psi_j(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)}\Big) \\
&\times \Big(\big|\big|xw_i(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)} + \big|\big|w_i(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)}\Big) \\
&+ \Big(\big|\big|x\psi_i(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)} + \big|\big|\psi_i(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)}\Big) \\
&\times \Big(\big|\big|xw_j(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)} + \big|\big|w_j(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)}\Big) \\
&+ \Big(\big|\big|xw_i(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)} + \big|\big|w_i(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)}\Big) \\
&\times \Big(\big|\big|xw_j(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)} + \big|\big|w_j(\tilde{t},x))\big|\big|_{L^2(\mathbb{R}^n)}\Big) \\
&= C(||\psi_i(\tilde{t},x)||_\Sigma)||w_j(\tilde{t},x)||_\Sigma + C(||\psi_j(\tilde{t},x)||_\Sigma)||w_i(\tilde{t},x)||_\Sigma + ||w_i(\tilde{t},x)||_\Sigma||w_j(\tilde{t},x)||_\Sigma \\
&\leq C(\delta_i+\delta_j) + \delta_i\delta_j \leq C\delta_{ij},
\end{aligned}
$$

for $\delta_i > 0$, $\delta_j, > 0$ and $\delta_i < \delta_{ij}$, $\delta_j < \delta_{ij}$. Now, if we proceed as in the proof of the first part of the theorem we get also that

$$
\begin{aligned}
&\Big|Cov(\tilde{\psi}_1(t,x)+\tilde{\psi}_3(t,x),\tilde{\psi}_2(t,x)+\tilde{\psi}_4(t,x) \\
&-Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x))\Big| \leq \delta
\end{aligned}
$$

for $\delta < \frac{1}{4}\min(\delta_{12},\delta_{14},\delta_{32},\delta_{34})$ and hence the statement. This completes the proof of Theorem 3.2.5.

3.2.5 Proof of Theorem 3.2.6

In this subsection, we give a complete proof of Theorem 3.2.6. We first prove that we can choose the parameter ω such that the mass of each soliton is $M_\omega = 1$.

Lemma 3.2.38. *There exists $\omega > 0$ such that $\|Q_\omega\|_{L^2(\mathbb{R}^n)} = 1$.*

Remark 3.2.39. *This is the only place in which we use the condition $1 < p < 1 + \frac{4}{n}$. For all the other steps, we just need the existence and the properties of the solitons.*

Proof. Consider $\lambda = \lambda(\omega) > 0$ and $Q_\lambda = \lambda^\alpha Q(\lambda x)$ with $\alpha > 0$, Q solution of $\Delta Q + Q^p = Q$ and Q_ω solutions of $\Delta Q + Q^p = \omega Q$. If we plug inside this last equation the ansatz $Q_\lambda = \lambda^\alpha Q(\lambda x)$, we get

$$\lambda^{\alpha+2}\Delta Q + \lambda^{\alpha p} Q^p = \omega \lambda^\alpha Q.$$

Therefore $\omega = \lambda^2$ and $\alpha = \frac{2}{p-1}$. Hence $Q_{\lambda(\omega)} = Q_\omega = \omega^{\frac{1}{p-1}} Q(\omega^{\frac{1}{2}} x)$. Now, we compute the mass of Q_ω which gives

$$M_\omega^2 = \int_{\mathbb{R}^n} |Q_\omega|^2 = \omega^{\frac{2}{p-1} - \frac{n}{2}} M^2,$$

with $M^2 = \int_{\mathbb{R}^n} |Q|^2$ independent of ω. So, by solving $1 = \omega^{\frac{2}{p-1} - \frac{n}{2}} M^2$ we get the thesis. □

From now on, all the functions Q will be Q_ω such that $M_\omega = 1$.

Remark 3.2.40. *This requirement becomes necessary in the following for the computation of the* Measure of Amalgamation*. It basically asks $|Q|^2$ to be a probability distribution.*

Let $\omega_k > 0$, $v_k \in \mathbb{R}^n$ and $\gamma_k^0 \in \mathbb{R}$. Assume that for any $k \neq k'$, we have $v_k \neq v_{k'}$. Let

$$R_k(t, x) = Q_{\omega_k}(x - x_k^0 - v_k^0 t) e^{i[\frac{1}{2} v_k^0 \cdot x - \frac{t}{4}|v_k^0|^2 + \omega_k t + \gamma_k^0]}.$$

Now, we compute the *Center of Mass* of the solitons in a way similar to how we computed the *Center of Mass* in the case of the *Quantum Harmonic Oscillator*.

Lemma 3.2.41. *Consider Q_ω such that $M_\omega = 1$. Then $\mu_k(t) := \int_{\mathbb{R}^n} x|R_k(t,x)|^2 = x_k^0 + v_k t$.*

Proof.

$$\mu_k(t) = \int_{\mathbb{R}^n} x|R_k(t,x)|^2 = \int_{\mathbb{R}^n} x|Q_\omega(x - x_k^0 - v_k t)|^2 = \int_{\mathbb{R}^n} (y + x_k^0 + v_k t))|Q_\omega(y)|^2 \quad (3.33)$$

$$= \int_{\mathbb{R}^n} (x_k^0 + v_k t))|Q_\omega(y)|^2 = (x_k^0 + v_k t) M_\omega^2 = x_k^0 + v_k t, \quad (3.34)$$

by Lemma 3.2.38. □

Now, we are ready to compute the *Measure of Amalgamation* between two moving solitons R_k and R_j.

$$Cov(R_k(t), R_j(t)) = \Re \int_{\mathbb{R}^n} (x - \mu_k(t))(x - \mu_j(t)) Q_\omega(x - x_k^0 - v_k t) \times$$
$$\times Q_\omega(x - x_j^0 - v_j t) e^{i[\theta_k(t,x) - \theta_j(t,x)]}$$

with

$$\theta_j(t, x) = \frac{1}{2} v_j^0 \cdot x - \frac{t}{4} |v_j^0|^2 + \omega_j t + \gamma_j^0$$

and

$$\theta_k(t, x) = \frac{1}{2} v_k^0 \cdot x - \frac{t}{4} |v_k^0|^2 + \omega_k t + \gamma_k^0.$$

Now, we choose

$$\omega_k^0 = \omega_j^0 = \omega$$

(the ω such that $M_\omega = 1$ from Lemma 3.2.38) and

$$\gamma_k^0 = \gamma_j^0 = \gamma$$

so that

$$\theta_k(t, x) - \theta_j(t, x) = \frac{1}{2} x(v_k - v_j) - \frac{1}{4} t(|v_k|^2 - |v_j|^2).$$

Now, we use a trick to reorganize the integrand in order to find again a structure similar to the one that we obtained in the case of the *Quantum Harmonic Oscillator*. We focus on the exponential terms first. We multiply and divide the integrand of $Cov(R_k(t), R_j(t))$ by

$$e^{\frac{1}{4}|x - \mu_k(t)|^2 + \frac{1}{4}|x - \mu_k(t)|^2}$$

and get

$$Cov(R_k(t), R_j(t)) = \Re \int_{\mathbb{R}^n} (x - \mu_k(t))(x - \mu_j(t)) e^{\frac{1}{4}|x-\mu_k(t)|^2 + \frac{1}{4}|x-\mu_k(t)|^2} \times$$
$$\times Q_\omega(x - x_k^0 - v_k t) Q_\omega(x - x_j^0 - v_j t) e^{-\frac{1}{4}\left[|x-\mu_k(t)|^2 + |x-\mu_j(t)|^2 + 2ix\cdot(v_k - v_j) + it(|v_k|^2 - |v_j|^2)\right]}.$$

Now, we complete the square to get

$$-\frac{1}{4}\left[|x - \mu_k(t)|^2 + |x - \mu_j(t)|^2 + 2ix \cdot (v_k - v_j)\right] =$$
$$-\frac{1}{2}\left[|x|^2 - x \cdot (\mu_k(t) + \mu_j(t)) + \frac{1}{2}|\mu_k(t)|^2 + \frac{1}{2}|\mu_j(t)|^2 + ix \cdot (v_k - v_j)\right] =$$
$$-\frac{1}{2}\left[|x|^2 - x \cdot (\mu_k(t) + \mu_j(t) - i(v_k - v_j)) + \frac{1}{2}|\mu_k(t)|^2 + \frac{1}{2}|\mu_j(t)|^2\right] =$$
$$-\frac{1}{4}\left[|\mu_k(t)|^2 + |\mu_j(t)|^2\right]$$
$$-\frac{1}{2}\left[|x|^2 - x \cdot (\mu_k(t) + \mu_j(t) - i(v_k - v_j)) + \frac{1}{4}((\mu_k(t) + \mu_j(t) - i(v_k - v_j))^2)\right]$$
$$+\frac{1}{8}(\mu_k(t) + \mu_j(t) - i(v_k - v_j))^2 =$$
$$-\frac{1}{4}\left[|\mu_k(t)|^2 + |\mu_j(t)|^2\right] + \frac{1}{8}(\mu_k(t) + \mu_j(t) - i(v_k - v_j))^2$$
$$-\frac{1}{2}\left[x - \frac{1}{2}(\mu_k(t) + \mu_j(t) - i(v_k - v_j))\right]^2$$

Now, putting everything back together and taking the change of variables $y := x - \frac{1}{2}(\mu_k(t) + \mu_j(t) - i(v_k - v_j))$, we get:

$$
\begin{aligned}
&Cov(R_k(t),R_j(t)) = \Re e^{-i\frac{t}{4}[|v_k|^2-|v_j|^2]}e^{-\frac{1}{4}\left[|\mu_k(t)|^2+|\mu_j(t)|^2\right]+\frac{1}{8}(\mu_k(t)+\mu_j(t)-i(v_k-v_j))^2}\times\\
&\int_{\mathbb{R}^n}\left(y+\frac{1}{2}(\mu_k(t)+\mu_j(t)-i(v_k-v_j))-\mu_k(t)\right)\\
&\times\left(y+\frac{1}{2}(\mu_k(t)+\mu_j(t)-i(v_k-v_j))-\mu_j(t)\right)\\
&\times e^{\frac{1}{4}|y+\frac{1}{2}(\mu_k(t)+\mu_j(t)-i(v_k-v_j))-\mu_j(t)|^2+\frac{1}{4}|y+\frac{1}{2}(\mu_k(t)+\mu_j(t)-i(v_k-v_j))-\mu_k(t)|^2}\times\\
&\times Q_\omega(x-x_k^0-v_kt)Q_\omega(x-x_j^0-v_jt)e^{-\frac{1}{2}|y|^2}=\\
&\Re e^{-i\frac{t}{4}[|v_k|^2-|v_j|^2]}e^{-\frac{1}{4}\left[|\mu_k(t)|^2+|\mu_j(t)|^2\right]+\frac{1}{8}(\mu_k(t)+\mu_j(t)-i(v_k-v_j))^2}\times\\
&\times\int_{\mathbb{R}^n}\left(y+\frac{1}{2}\left(\mu_k(t)-\mu_j(t)-i(v_k-v_j)\right)\right)\\
&\times\left(y+\frac{1}{2}\left(-\mu_k(t)+\mu_j(t)-i(v_k-v_j)\right)\right)\times\\
&\times e^{\frac{1}{4}|x-\mu_j(t)|^2+\frac{1}{4}|x-\mu_k(t)|^2}\times Q_\omega(x-x_k^0-v_kt)Q_\omega(x-x_j^0-v_jt)e^{-\frac{1}{2}|y|^2}=\\
&\Re e^{-i\frac{t}{4}[|v_k|^2-|v_j|^2]}e^{-\frac{1}{4}\left[|\mu_k(t)|^2+|\mu_j(t)|^2\right]+\frac{1}{8}(\mu_k(t)+\mu_j(t)-i(v_k-v_j))^2}\times\\
&\times\int_{\mathbb{R}^n}\left[(y-i\frac{1}{2}(v_k-v_j))^2-\frac{1}{4}|\mu_k(t)-\mu_j(t)|^2\right]\times\\
&\times e^{\frac{1}{4}|x-\mu_j(t)|^2+\frac{1}{4}|x-\mu_k(t)|^2}\times Q_\omega(x-x_k^0-v_kt)Q_\omega(x-x_j^0-v_jt)e^{-\frac{1}{2}|y|^2}=\\
&\Re e^{-i\frac{t}{4}[|v_k|^2-|v_j|^2]}e^{-\frac{1}{4}\left[|\mu_k(t)|^2+|\mu_j(t)|^2\right]+\frac{1}{8}(\mu_k(t)+\mu_j(t)-i(v_k-v_j))^2}\times\\
&\times\int_{\mathbb{R}^n}\left[|y|^2-i(v_k-v_j)y-\frac{1}{4}|v_k-v_j|^2-\frac{1}{4}|\mu_k(t)-\mu_j(t)|^2\right]\times\\
&\times e^{\frac{1}{4}|x-\mu_j(t)|^2+\frac{1}{4}|x-\mu_k(t)|^2}\times Q_\omega(x-x_k^0-v_kt)Q_\omega(x-x_j^0-v_jt)e^{-\frac{1}{2}|y|^2}.
\end{aligned}
$$

Now, we need to get rid of the odd term in y inside the integral and, for this, we need the following lemma.

Lemma 3.2.42. *For every $p\in\mathbb{R}^n$ we have $\int_{\mathbb{R}^n} yQ(y-p)Q(y+p)dy=0$.*

Proof.

$$
\begin{aligned}
&\int_{\mathbb{R}^n} yQ(y-p)Q(y+p)dy=\int_{\mathbb{R}^n}-zQ(-z-p)Q(-z+p)dz\\
&=-\int_{\mathbb{R}^n} zQ(z+p)Q(z-p)dz=-\int_{\mathbb{R}^n} yQ(y-p)Q(y+p)dy=0
\end{aligned}
$$

by radiality of Q. □

Therefore, going on with the computation and using again the radiality of Q_ω, we get

$$
\begin{aligned}
&Cov(R_k(t), R_j(t)) = \Re e^{-i\frac{t}{4}[|v_k|^2 - |v_j|^2]} e^{-\frac{1}{4}\left[|\mu_k(t)|^2 + |\mu_j(t)|^2\right] + \frac{1}{8}(\mu_k(t)+\mu_j(t) - i(v_k - v_j))^2} \\
&\times \int_{\mathbb{R}^n} \left[|y|^2 - \frac{1}{4}|v_k - v_j|^2 - \frac{1}{4}|\mu_k(t) - \mu_j(t)|^2 \right] \\
&\times e^{\frac{1}{4}|x-\mu_j(t)|^2 + \frac{1}{4}|x - \mu_k(t)|^2} \times Q_\omega(x - x_k^0 - v_k t) Q_\omega(x - x_j^0 - v_j t) e^{-\frac{1}{2}|y|^2} \\
&= \cos\left(-\frac{t}{4}[|v_k|^2 - |v_j|^2] - \frac{1}{4}(\mu_k(t) + \mu_j(t))(v_k - v_j) \right) \\
&\times e^{-\frac{1}{4}\left[|\mu_k(t)|^2 + |\mu_j(t)|^2\right] + \frac{1}{8}|\mu_k(t) + \mu_j(t)|^2 - \frac{1}{8}|v_k - v_j|^2} \\
&\times \int_{\mathbb{R}^n} \left[|y|^2 - \frac{1}{4}|v_k - v_j|^2 - \frac{1}{4}|\mu_k(t) - \mu_j(t)|^2 \right] \\
&\times e^{\frac{1}{4}|y + \frac{1}{2}(\mu_k(t) - \mu_j(t) - i(v_k - v_j)|^2 + \frac{1}{4}|y - \frac{1}{2}(\mu_k(t) - \mu_j(t) + i(v_k - v_j)|^2} \\
&\times Q_\omega\left(\left| y + \frac{1}{2}(\mu_k(t) - \mu_j(t) - i(v_k - v_j) \right| \right) \times \\
&Q_\omega\left(\left| y - \frac{1}{2}(\mu_k(t) - \mu_j(t) + i(v_k - v_j) \right| \right) e^{-\frac{1}{2}|y|^2} \\
&= \cos\left(-\frac{t}{4}[|v_k|^2 - |v_j|^2] - \frac{1}{4}(\mu_k(t) + \mu_j(t))(v_k - v_j) \right) \times \\
&\times e^{+\frac{1}{8}|\mu_k(t) - \mu_j(t)|^2 + \frac{1}{8}|v_k - v_j|^2} \\
&\times \int_{\mathbb{R}^n} \left[|y|^2 - \frac{1}{4}|v_k - v_j|^2 - \frac{1}{4}|\mu_k(t) - \mu_j(t)|^2 \right] \times \\
&\times e^{-\frac{1}{8}|\mu_k(t) - \mu_j(t)|^2 - \frac{1}{8}|v_k - v_j|^2} \\
&\times Q_\omega\left(\left| y + \frac{1}{2}(\mu_k(t) - \mu_j(t) - i(v_k - v_j) \right| \right) \\
&\times Q_\omega\left(\left| y - \frac{1}{2}(\mu_k(t) - \mu_j(t) + i(v_k - v_j) \right| \right) =
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
&Cov(R_k(t), R_j(t)) = \\
&\cos\left(\frac{t}{4}[|v_k|^2 - |v_j|^2] + \frac{1}{4}(\mu_k(t) + \mu_j(t))(v_k - v_j) \right) \\
&\times \int_{\mathbb{R}^n} \left[|y|^2 - \frac{1}{4}|v_k - v_j|^2 - \frac{1}{4}|\mu_k(t) - \mu_j(t)|^2 \right] \\
&\times Q_\omega\left(\left| y + \frac{1}{2}\left(\mu_k(t) - \mu_j(t) - i(v_k - v_j)\right) \right| \right) \\
&\times Q_\omega\left(\left| y - \frac{1}{2}\left(\mu_k(t) - \mu_j(t) + i(v_k - v_j)\right) \right| \right).
\end{aligned}
$$

Now, we change variables again to $y = x - \frac{i}{2}(v_k - v_j)$ ad we get:

$$Cov(R_k(t), R_j(t)) = \cos\left(\frac{t}{4}[|v_k|^2 - |v_j|^2] + \frac{1}{4}(\mu_k(t) + \mu_j(t))(v_k - v_j)\right) \times$$
$$\times \int_{\mathbb{R}^n} \left[|y|^2 - \frac{1}{4}|\mu_k(t) - \mu_j(t)|^2\right]$$
$$\times Q_\omega\left(\left|y + \frac{1}{2}(\mu_k(t) - \mu_j(t))\right|\right) Q_\omega\left(\left|y - \frac{1}{2}(\mu_k(t) - \mu_j(t))\right|\right).$$

To enlighten the notation, we define $a_{kj} := \frac{1}{2}(\mu_k(t) - \mu_j(t))$ and hence we get:

$$Cov(R_k(t), R_j(t)) = \cos\left(\frac{t}{4}[|v_k|^2 - |v_j|^2] + \frac{1}{2}a_{kj}(v_k - v_j)\right) \times$$
$$\times \int_{\mathbb{R}^n} \left[|y|^2 - |a_{kj}|^2\right] Q_\omega(|y + a_{kj}|) Q_\omega(|y - a_{kj}|).$$

Remark 3.2.43. *Until now, we did not use the hypothesis $n = 1$. We will from now on.*

Now, we use the explicit form of the soliton:

$$Q(x) = \left(\frac{p+1}{2}\right)^{p-1}\left[\cosh\left(\frac{p-1}{2}x\right)\right]^{-2p+2}$$

up to symmetries (see Subsection 3.2.2). Therefore, we have

$$Q_\omega\left(\left|x + \frac{1}{2}(\mu_k(t) - \mu_j(t))\right|\right) Q_\omega\left(\left|x - \frac{1}{2}(\mu_k(t) - \mu_j(t))\right|\right) =$$
$$\left(\frac{p+1}{2}\right)^{p-1}\omega^{\frac{1}{p-1}}\left[\cosh\left(\frac{p-1}{2}\omega^{\frac{1}{2}}[x + a_{kj}]\right)\right]^{-2p+2} \times$$
$$\times\omega^{\frac{1}{p-1}}\left(\frac{p+1}{2}\right)^{p-1}\left[\cosh\left(\frac{p-1}{2}\omega^{\frac{1}{2}}[x - a_{kj}]\right)\right]^{-2p+2} =$$
$$\left(\frac{p+1}{2}\right)^{2p-2}\omega^{\frac{2}{p-1}}\left[\cosh\left(\frac{p-1}{2}\omega^{\frac{1}{2}}[x + a_{kj}]\right)\right]^{-2p+2}$$
$$\times\left[\cosh\left(\frac{p-1}{2}\omega^{\frac{1}{2}}[x - a_{kj}]\right)\right]^{-2p+2}.$$

Now, since $v_k \neq v_j$ for $j \neq k$ either $v_k > v_j$ or $v_k < v_j$. We suppose $v_k > v_j$. Therefore, for $t \gg 1$, we have $\mu_k(t) > \mu_j(t)$ and so as $t \to +\infty$, then $a_{kj} \to +\infty$. Therefore

$$Q_\omega \left(\left| x + \frac{1}{2} \left(\mu_k(t) - \mu_j(t) \right) \right| \right) Q_\omega \left(\left| y - \frac{1}{2} \left(\mu_k(t) - \mu_j(t) \right) \right| \right) =$$
$$\left(\frac{p+1}{2} \right)^{2p-2} \omega^{\frac{2}{p-1}} \left[\cosh \left(\frac{p-1}{2} \omega^{\frac{1}{2}} [x + a_{kj}] \right) \right]^{-2p+2}$$
$$\times \left[\cosh \left(\frac{p-1}{2} \omega^{\frac{1}{2}} [x - a_{kj}] \right) \right]^{-2p+2} \simeq$$
$$(2(p+1))^{2p-2} \omega^{\frac{2}{p-1}} e^{-(p-1)^2 \omega^{\frac{1}{2}} \left[|x+a_{kj}| + |x-a_{kj}| \right]},$$

uniformly in x over $\mathbb{R}$, as $a_{kj} \to +\infty$. Therefore,

$$\int_{\mathbb{R}^n} \left[|y|^2 - |a_{kj}|^2 \right] Q_\omega \left(|y + a_{kj}| \right) Q_\omega \left(|y - a_{kj}| \right) \simeq$$
$$(2(p+1))^{2p-2} \omega^{\frac{2}{p-1}} \int_{\mathbb{R}} \left[|x|^2 - |a_{kj}|^2 \right] e^{-(p-1)^2 \omega^{\frac{1}{2}} \left[|x+a_{kj}| + |x-a_{kj}| \right]}.$$

We compute this last integral explicitly, dividing it into three parts:

$$\int_{\mathbb{R}} \left[|x|^2 - |a_{kj}|^2 \right] e^{-(p-1)^2 \omega^{\frac{1}{2}} \left[|x+a_{kj}| + |x-a_{kj}| \right]} =$$
$$\int_{x > a_{kj}} \left[|x|^2 - |a_{kj}|^2 \right] e^{-(p-1)^2 \omega^{\frac{1}{2}} \left[|x+a_{kj}| + |x-a_{kj}| \right]} +$$
$$\int_{-a_{kj} < x < a_{kj}} \left[|x|^2 - |a_{kj}|^2 \right] e^{-(p-1)^2 \omega^{\frac{1}{2}} \left[|x+a_{kj}| + |x-a_{kj}| \right]} +$$
$$\int_{x < -a_{kj}} \left[|x|^2 - |a_{kj}|^2 \right] e^{-(p-1)^2 \omega^{\frac{1}{2}} \left[|x+a_{kj}| + |x-a_{kj}| \right]} = I(a_{kj}) + II(a_{kj}) + III(a_{kj}).$$

Now, we compute each integral separately and then we put everything together. We start with $III(a_{kj})$:

$$III(a_{kj}) := \int_{x < -a_{kj}} \left[|x|^2 - |a_{kj}|^2 \right] e^{-(p-1)^2 \omega^{\frac{1}{2}} \left[|x+a_{kj}| + |x-a_{kj}| \right]} dx =$$
$$- \int_{-z < -a_{kj}} \left[|-z|^2 - |a_{kj}|^2 \right] e^{-(p-1)^2 \omega^{\frac{1}{2}} \left[|-z+a_{kj}| + |-z-a_{kj}| \right]} - (dz) =$$
$$\int_{z > a_{kj}} \left[|z|^2 - |a_{kj}|^2 \right] e^{-(p-1)^2 \omega^{\frac{1}{2}} \left[|z-a_{kj}| + |+z+a_{kj}| \right]} dz = I(a_{kj}).$$

Now, assuming $a_{kj} > 0$ (but for $a_{kj} < 0$ it works in the same way), we have

$$I(a_{kj}) = \int_{x>a_{kj}} \left[|x|^2 - |a_{kj}|^2\right] e^{-(p-1)^2\omega^{\frac{1}{2}}\left[|x+a_{kj}|+|x-a_{kj}|\right]} =$$
$$\int_{x>a_{kj}} \left[|x|^2 - |a_{kj}|^2\right] e^{-2(p-1)^2\omega^{\frac{1}{2}}x} =$$
$$\frac{1}{2}(p-1)^{-2}\omega^{-\frac{1}{2}} \int_{y>2(p-1)^2\omega^{\frac{1}{2}}a_{kj}} \left[\frac{1}{4}(p-1)^{-4}\omega^{-1}|y|^2 - |a_{kj}|^2\right] e^{-y} =$$
$$\frac{1}{8(p-1)^6}\omega^{-\frac{3}{2}} \left\{\left[2(p-1)^2\omega^{\frac{1}{2}}a_{kj} + 1\right]^2 + 1\right\}$$
$$\times e^{-2(p-1)^2\omega^{\frac{1}{2}}a_{kj}} - \frac{|a_{kj}|^2\omega^{-\frac{1}{2}}}{2(p-1)^2} e^{-\frac{1}{2}(p-1)^2\omega^{\frac{1}{2}}a_{kj}}.$$

by changing variables $y := 2(p-1)^2\omega^{\frac{1}{2}}x$. Also,

$$II(a_{kj}) := \int_{-a_{kj}<x<a_{kj}} \left[|x|^2 - |a_{kj}|^2\right] e^{-(p-1)^2\omega^{\frac{1}{2}}\left[|x+a_{kj}|+|x-a_{kj}|\right]} =$$
$$\int_{-a_{kj}<x<a_{kj}} \left[|x|^2 - |a_{kj}|^2\right] e^{-2a_{kj}(p-1)^2\omega^{\frac{1}{2}}} = e^{-2a_{kj}(p-1)^2\omega^{\frac{1}{2}}} \left[\frac{x^3}{3} - a_{kj}^2 x\right]_{-a_{kj}}^{+a_{kj}} =$$
$$-\frac{4}{3}a_{kj}^3 e^{-2a_{kj}(p-1)^2\omega^{\frac{1}{2}}}.$$

Therefore,

$$I(a_{kj}) + II(a_{kj}) + III(a_{kj}) \simeq -\frac{4}{3}a_{kj}^3 e^{-2a_{kj}(p-1)^2\omega^{\frac{1}{2}}},$$

when $t \to +\infty$, since $II(a_{kj})$ has a stronger power of a_{kj}. This implies that

$$\int_{\mathbb{R}^n} \left[|y|^2 - |a_{kj}|^2\right] Q_\omega\left(|y + a_{kj}|\right) Q_\omega\left(|y - a_{kj}|\right) \longrightarrow 0^-$$

as $t \to +\infty$ and hence it is asymptotically negative. Now, we concentrate on the phase:

$$\cos\left(\frac{t}{4}[|v_k|^2 - |v_j|^2] + \frac{1}{2}a_{kj}(v_k - v_j)\right)$$
$$\times = \cos\left(\frac{t}{2}[|v_k|^2 - |v_j|^2] + \frac{1}{4}(x_0^k + x_0^j)(v_k - v_j)\right).$$

We choose parameters in the following way:

$$v_1 = -v_2, \quad v_3 = -v_4, \quad \frac{1}{4}(x_0^1 + x_0^2)(v_1 - v_2) = \frac{\pi}{2} + \epsilon, \text{ and } \frac{1}{4}(x_0^3 + x_0^4)(v_3 - v_4) = -\frac{3}{2}\pi + \delta,$$

so that

$$Cov(R_1(t), R_2(t)) \simeq \frac{4}{3}\epsilon a_{12}^3 \left(\frac{p+1}{2}\right)^{2p-2} \omega^{\frac{2}{p-1}} e^{-2a_{12}(p-1)^2\omega^{\frac{1}{2}}}$$

and

$$Cov(R_3(t), R_4(t)) \simeq \frac{4}{3}\delta a_{34}^3 \left(\frac{p+1}{2}\right)^{2p-2} \omega^{\frac{2}{p-1}} e^{-2a_{34}(p-1)^2\omega^{\frac{1}{2}}},$$

for $0 < \epsilon \ll 1$ and $0 < \delta \ll 1$.

Moreover, we choose $x_0^2 = -x_0^3$ and $x_0^1 = -x_0^4$, $|v_4|^2 - |v_1|^2 = q_1$, $|v_3|^2 - |v_2|^2 = q_2$, $\tilde{t}_1$ such that $\frac{1}{2}\tilde{t}_1 q_2 = 0 \bmod 2\pi$ and $\tilde{t}_2$ such that $\frac{1}{2}\tilde{t}_2 q_1 = 0 \bmod 2\pi$. These choices imply

$$q_1 = q_2 = q, \quad (v_1, v_2, v_3, v_4) = (v_1, -v_1, +\sqrt{q + |v_1|^2}, -\sqrt{q + |v_1|^2})$$

and

$$\tilde{t} = \tilde{t}_1 = \tilde{t}_2 = \frac{4\pi k}{q}$$

with $k \gg 1$ (basically big enough to satisfy the asymptotic conditions of Theorem 3.2.10 and [88]), v_1 such that

$$|v_1|^2 = q \frac{\left(\frac{\pi}{2} + \epsilon\right)^2}{-\left(\frac{\pi}{2} + \epsilon\right)^2 + \left(-\frac{3}{2}\pi + \delta\right)^2}$$

and

$$(x_0^1, x_0^2, x_0^3, x_0^4) = \left(x_0^1, \frac{2}{v}\left(\frac{\pi}{2} + \epsilon\right) - x_0^1, -\frac{2}{v}\left(\frac{\pi}{2} + \epsilon\right) + x_0^1, -x_0^1\right)$$

with $v = v_1$.

At this point, we can compute

$$\begin{aligned}
&Cov(R_1(\tilde{t},x)+R_3(\tilde{t},x),R_2(\tilde{t},x)+R_4(\tilde{t},x)) = \\
&Cov(R_1(\tilde{t},x),R_2(\tilde{t},x))+Cov(R_3(\tilde{t},x),R_4(\tilde{t},x))+ \\
&Cov(R_2(\tilde{t},x),R_3(\tilde{t},x))+Cov(R_4(\tilde{t},x),R_1(\tilde{t},x)) \simeq \\
&\frac{4}{3}\epsilon a_{12}^3\left(\frac{p+1}{2}\right)^{2p-2}\omega^{\frac{2}{p-1}}e^{-2a_{12}(p-1)^2\omega^{\frac{1}{2}}}+\frac{4}{3}\delta a_{34}^3\left(\frac{p+1}{2}\right)^{2p-2}\omega^{\frac{2}{p-1}}e^{-2a_{34}(p-1)^2\omega^{\frac{1}{2}}} \\
&-\frac{4}{3}a_{23}^3\left(\frac{p+1}{2}\right)^{2p-2}\omega^{\frac{2}{p-1}}e^{-2a_{23}(p-1)^2\omega^{\frac{1}{2}}}-\frac{4}{3}a_{41}^3\left(\frac{p+1}{2}\right)^{2p-2}\omega^{\frac{2}{p-1}}e^{-2a_{41}(p-1)^2\omega^{\frac{1}{2}}}<0,
\end{aligned}$$

for ϵ and δ small enough (similarly as in the proof for the *Quantum Harmonic Oscillator*). Moreover, since $\tilde{t} \gg 1$ and from Theorem 3.2.10, one has that $\psi(\tilde{t},x) \simeq \Sigma_{k=1}^4 R_k(\tilde{t},x)$ and so, at the instant $\tilde{t} \gg 1$, the *Simpson's Paradox* occurs. This completes the proof of Theorem 3.2.6.

3.2.6 How likely is the Simpson's Paradox in Quantum Mechanics?

An important question is: "How likely is the Simpson's Paradox?". It is in fact interesting to quantify, in some way, the chances that one has to run into the paradox.

In the case of $2 \times 2 \times l$ contingency tables with $l \geq 2$, Pavlides and Perlman [94] address the problem and, among the other things, they prove the following.

Suppose that a contingency table consists of a factor A with two levels, a factor B with other 2 levels and a third factor C with $l \geq 2$-levels. Then, the array of cell probabilities $\mathbf{p}$ lies on the Simplex

$$\mathcal{S}_{4l} := \left\{\mathbf{p} |\ p_i \geq 0,\ \forall i = 1, \ldots, 4l;\ \Sigma_{i=1}^{4l} p_i = 1\right\}.$$

Endow $\mathcal{S}_{4l}$ with the *Dirichlet Distribution on* $\mathcal{S}_{4l}$, denoted by $D_{4l}(\alpha)$ and denote with $\pi_l(\alpha)$ the probability of having the *Simpson's Paradox* under $D_{4l}(\alpha)$. Pavlides and Perlman proved in [94] that $\pi_2(1) = \frac{1}{60}$ and conjectured that for every $\alpha > 0$, there exists $h(\alpha) > 0$ such that

$$\pi_l(\alpha) \simeq \pi_2(\alpha) \times e^{-h(\alpha)\left(\frac{l}{2}-1\right)}, \quad l = 2, 3, \ldots.$$

A similar question can be asked in the case of the *Quantum Harmonic Oscillator* and the *Nonlinear Schrödinger equation.* In the constructions developed in the previous sections, we aimed just at finding one single choice of the parameters which gives the *Simpson's Paradox* and we did it mainly with a perturbative method. But how large is (and in which sense it is large) the set of parameters which gives the *Simpson's Paradox*?

To investigate a little bit further this issue, we briefly recall the proof of Theorem 3.2.1, at least in the stationary case and deduce from it a preliminary result on the likelihood of occurrence of the *Simpson's Paradox.*

Consider two *moving solitons* of the form:

$$\psi_i(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_i(t)+\gamma_i(t)+\frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x-x_i(t)|^2},$$

and

$$\psi_j(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_j(t)+\gamma_j(t)+\frac{\omega t}{2}]} e^{-\frac{m\omega}{2\hbar}|x-x_j(t)|^2},$$

for $1 \leq i \leq j \leq 4$ and with $\mathbf{x}(t)$, $\mathbf{v}(t)$ and $\gamma(t)$ as in Subsection 3.2.8.

Consider the case in which, for every $t \in \mathbb{R}$, one has that $x_k(t) = x_k$, for every $k = 1, \ldots, N$ independent of time. It has been proven in Proposition 3.2.16 above that the *Covariance* between any of these two solitons is given by:

$$Cov(\psi_i(t,x), \psi_j(t,x)) = \frac{1}{2}\cos(\gamma_i - \gamma_j)\left[\frac{\hbar}{m\omega} - \frac{1}{2}|x_i - x_j|^2\right] e^{-\frac{m\omega}{4\hbar}|x_i-x_j|^2}$$

Therefore, the proof of Theorem 3.2.1 in the stationary case reduces to the problem of finding parameters such that the *Simpson's Paradox* occurs, namely such that

$$Cov(\psi_1(t,x), \psi_2(t,x)) > 0,$$

$$Cov(\psi_3(t,x), \psi_4(t,x)) > 0$$

but

$$Cov(\psi_1(t,x) + \psi_3(t,x), \psi_2(t,x) + \psi_4(t,x)) < 0$$

or vice versa,

$$Cov(\psi_1(t,x),\psi_2(t,x)) < 0,$$

$$Cov(\psi_3(t,x),\psi_4(t,x)) < 0$$

but

$$Cov(\psi_1(t,x)+\psi_3(t,x),\psi_2(t,x)+\psi_4(t,x)) > 0.$$

Now, we define

$$L_{ij}^2 := \frac{m\omega}{2\hbar}|x_i - x_j|^2$$

so that $Cov(\psi_i(t,x),\psi_j(t,x))$ can be rewritten in the following way:

$$Cov(\psi_i(t,x),\psi_j(t,x)) = \frac{\hbar}{2m\omega}\cos(\gamma_i - \gamma_j)\left[1 - L_{ij}^2\right] e^{-\frac{1}{2}L_{ij}^2}.$$

In the following discussion, we treat only the case $\gamma_i = \gamma_j$, for every $i, j = 1, \ldots, 4$.

We can restate our hypotheses and thesis in the following way: we suppose that $0 < L_{12} < 1$ and $0 < L_{34} < 1$ and we want to quantify "how many" admissible choices of $0 < L_{12} < 1$ and $0 < L_{34} < 1$, L_{23} and L_{14} there are such that

$$\left[1 - L_{12}^2\right] e^{-\frac{1}{2}L_{12}^2} + \left[1 - L_{23}^2\right] e^{-\frac{1}{2}L_{23}^2} + \left[1 - L_{34}^2\right] e^{-\frac{1}{2}L_{34}^2} + \left[1 - L_{14}^2\right] e^{-\frac{1}{2}L_{14}^2} < 0.$$

Remark 3.2.44. *Note that the defining condition for the occurrence of the* Simpson's Paradox *are all inequalities which is a hint of the fact that the* Simpson's Paradox *occurs in a open set of the correct topology (see Theorem 3.2.5).*

Since we are in dimension $n = 1$, we can choose $x_1 < x_2 < x_3 < x_4$. This implies that $L_{14} = L_{12} + L_{23} + L_{34}$ and so that we have to find an admissible choice of $0 < L_{12} < 1$ and $0 < L_{34} < 1$ and L_{23} such that

$$\begin{aligned}&\left[1 - L_{12}^2\right] e^{-\frac{1}{2}L_{12}^2} + \left[1 - L_{23}^2\right] e^{-\frac{1}{2}L_{23}^2} + \\ &+ \left[1 - L_{34}^2\right] e^{-\frac{1}{2}L_{34}^2} + \left[1 - (L_{12} + L_{23} + L_{34})^2\right] e^{-\frac{1}{2}(L_{12}+L_{23}+L_{34})^2} < 0.\end{aligned}$$

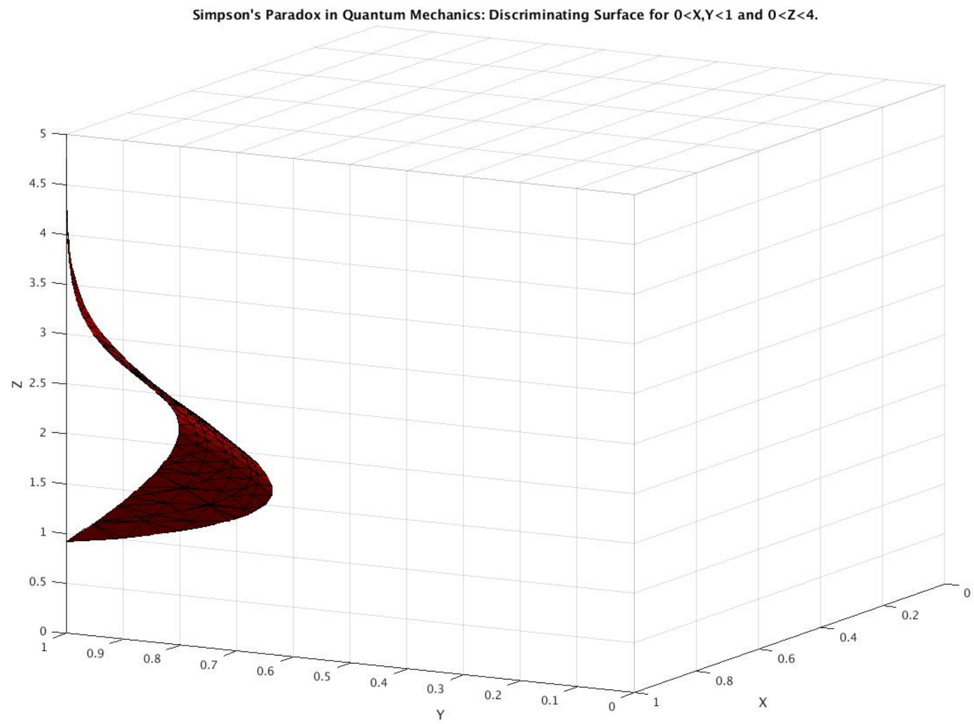


Figure 3.1: Surface discriminating between the region of parameters where the Simpson's Paradox occurs and does not occur.

Now, if we define $X := L_{12}$, $Y := L_{34}$ and $Z := L_{23}$, we get that the *Simpson's Paradox* occurs when the following are satisfied:

$$\begin{cases} 0 < X < 1 \\ 0 < Y < 1 \\ \left[1 - X^2\right] e^{-\frac{1}{2}X^2} + \left[1 - Y^2\right] e^{-\frac{1}{2}Y^2} + \left[1 - Z^2\right] e^{-\frac{1}{2}Z^2} + \\ \left[1 - (X + Y + Z)^2\right] e^{-\frac{1}{2}(X+Y+Z)^2} < 0. \end{cases}$$

Figure 1 focuses on a small region of the parameters' space with $0 < X, Y < 1$ and represents the surface which discriminates between where the paradox occurs and when it does not.

Note that, when one of the coordinates (for example Z) becomes larger and larger, the

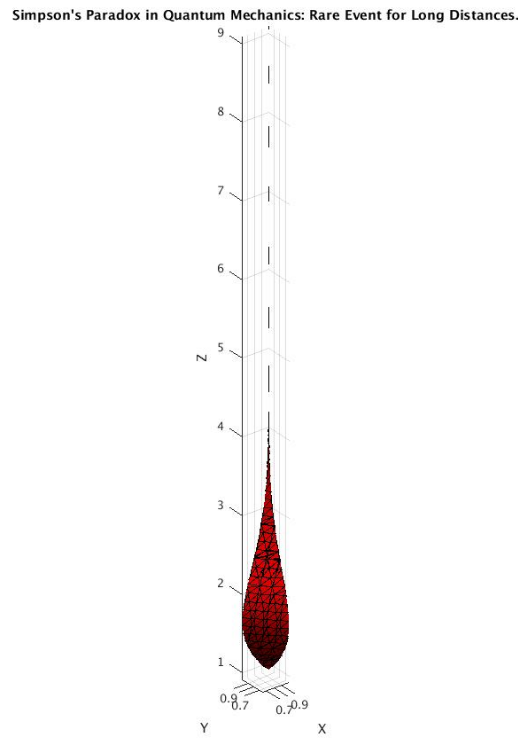


Figure 3.2: The likelihood of occurrence of the *Simpson's Paradox* decreases to zero as one of the distances between the particles increases indefinitely.

paradox occurs more and more rarely. In fact, the condition

$$\left[1-X^2\right]e^{-\frac{1}{2}X^2}+\left[1-Y^2\right]e^{-\frac{1}{2}Y^2}+\left[1-Z^2\right]e^{-\frac{1}{2}Z^2}+\left[1-(X+Y+Z)^2\right]e^{-\frac{1}{2}(X+Y+Z)^2}<0$$

for big Z reduces to

$$\left[1-X^2\right]e^{-\frac{1}{2}X^2}+\left[1-Y^2\right]e^{-\frac{1}{2}Y^2}<0$$

which is incompatible with

$$0<X<1,\ \textit{and}\ 0<Y<1.$$

Figure 2 explains this last sentence visually.

We have decided to test the inequality $f(X,Y,Z)<0$ over a grid of $n\times n\times n$ values with $n=1000$ in the parallelepiped $(X,Y,Z)\in[0,1]\times[0,1]\times[0,4]$ and we discovered that about $1.2*10^{-4}$ of the times (0.012%) the inequality is satisfied. Note that the choice of the uniform distribution on $[0,1]\times[0,1]\times[0,4]$ has been made because for $Z > 4$ the *Simpson's Paradox*'s region is almost null (Figure 2) and because already $0 < X, Y < 1$. This result deserves further investigation. For reproducibility purposes, we give the Matlab Code that we used for the analysis:

```
syms X Y Z
fun=@(X,Y,Z)((1-X.^2).*exp(-X.^2/2)+(1-Y.^2).*exp(-Y.^2/2)
+(1-Z.^2).*exp(-Z.^2/2)+(1-(X+Y+Z).^2).*exp(-(X+Y+Z).^2/2));

n=1000;
S=zeros(n);
m=zeros(n);
x=0:1/n:1;
y=0:1/n:1;
%since the max of this function is 4
z=0:1/n:4;
SP=0;
%syms t
%[X,Y]=meshgrid(0:0.1:1,0:0.1:1);
for i= 1:n+1
    for j=1:n+1
        for k=1:n+1
            if fun(x(i),y(j),z(k))<0
     SP=SP+1;
            else
           SP=SP+0;
    end
    end
end
SP # Number of occasions in which the Simpson's Paradox occurs
SP/(n+1)^3
# Percentage of occasions in which the Simpson's Paradox occurs
```

3.2.7 Some Numerical Examples

For illustration purposes, we give some numerical examples of cases in which the *Simpson's Paradox* occurs and on which it does not. We find interesting to give to each parameters their true physical value.

Consider the *Planck Constant*

$$\hbar = \frac{h}{2\pi} = \frac{1}{2\pi} * 6.62607004 * 10^{-34} m^2 kg/s = 1.0545718 * 10^{-34} m^2 kg/s,$$

the *Mass of an Electron*

$$m = 9.10938356 * 10^{-31} kg$$

with frequency of revolution

$$f = 6.6 * 10^{15} s^{-1}$$

and angular velocity

$$\omega = 2\pi f = 4.1469023 * 10^{16} s^{-1}.$$

Note that the quantity

$$L_{ij}^2 := \frac{m\omega}{2\hbar} |x_i - x_j|^2$$

that we defined and used in Section 3.2.6 for the sketch of the proof of the stationary case of Theorem 3.2.1, is dimensionless and it is a fundamental quantity.

We choose L_{12}^2,L_{34}^2 and L_{23}^2 which are all around 1. Note that this implies the following about the distance between the particles:

$$1 \simeq L_{ij}^2 = \frac{m\omega}{2\hbar} |x_i - x_j|^2 = \frac{9.10938356 * 10^{-31} * 4.1469023 * 10^{16}}{1.0545718 * 10^{-34}} |x_i - x_j|^2$$

$$\simeq 3.582091 * 10^{20} |x_i - x_j|^2.$$

This implies that

$$|x_i - x_j| \simeq 5.2836213 * 10^{-11} m.$$

Recall that the *Bohr Radius*, which represents approximately the most probable distance between the center of a nuclide and the electron in a hydrogen atom in its ground state, is

$$r_{Bohr} = 5.2917721067 * 10^{-11} m$$

We choose $L_{12}^2 = 1 - \epsilon_1^2$, $L_{34}^2 = 1 - \epsilon_2^2$ and $L_{23}^2 = 1 + \delta^2$ with $\epsilon_1 \ll 1$, $\epsilon_2 \ll 1$. The following R code produces and example of the paradox in our case:

```
x=1-10^(-10); #L_{12}^2<1--> Positive Correlation
```

```
y=1-10^(-10); #L_{34}^2<1--> Positive Correlation
z=1+10^(-5); #L_{23}^2
(1-x^2)*exp(-x^2/2)+(1-y^2)*exp(-y^2/2)+(1-z^2)*exp(-z^2/2)
+(1-(x+y+z)^2)*exp(-(x+y+z)^2/2)
#Reversal Condition <--> Negative Correlation
[1] -0.0888821
```

Of course, there are cases in which the *Simpson's Paradox* does not occur, like

```
x=1-10^(-1); #L_{12}^2<1--> Positive Correlation
y=1-10^(-1); #L_{34}^2<1--> Positive Correlation
z=1+10^(-5); #L_{23}^2
(1-x^2)*exp(-x^2/2)+(1-y^2)*exp(-y^2/2)+(1-z^2)*exp(-z^2/2)
+(1-(x+y+z)^2)*exp(-(x+y+z)^2/2)
#Reversal Condition not satisfied <--> Positive Correlation
[1] 0.1177287
```

3.2.8 Final Considerations and Open Problems

In this subsection, we present some open problems and some final considerations about the occurrence of the *Simpson's Paradox* in the settings of the *Quantum Harmonic Oscillator* and of the *Nonlinear Schrödinger Equation*.

• *Intermittent Paradox in the nonlinear case.* In the nonlinear case the choice of the parameters can be redone in order to get the non occurrence of the paradox. For example, keeping the same other conditions, but now requesting

$$\frac{1}{2}\tilde{t}_1 q_2 = \pi \text{ mod } 2\pi$$

and

$$\frac{1}{2}\tilde{t}_2 q_1 = \pi \text{ mod } 2\pi,$$

we get that the paradox disappears at time $\tilde{t} = \frac{2\pi}{q_1} mcm\{\frac{q_2}{2\pi}\tilde{t}_1, \frac{q_1}{2\pi}\tilde{t}_2\}$. Therefore, we can say that, for the *Nonlinear Schrödinger Equation*, the *Simpson's Paradox* is *Intermittent.* We think that this is the case for any choice of the initial data, but this is not proved in our theorems and so it remains an open problem.

• *What happens for finite times in the nonlinear case?.* This is hard to tell, because one does not have the soliton structure, which, in the nonlinear case, appears just asymptotically. We believe that, if one defines, in a reasonable way, a *measure of amalgamation* for solutions which are not solitons, would get that, for finite times, the *Simpson's Paradox* will be highly intermittent. In that case it would be interesting to understand how many times the paradox appears and disappears before a fixed time T.

• *The nonlinear case in spatial dimension $n > 1$.* In dimension $n > 1$, the explicit form of the ground state is not available. Hence the analysis done do not apply directly. Anyways, we still believe that, since the structure of the *measure of amalgamation* is similar to the other cases (as shown in Section 3.2.5 in the proof of Theorem 3.2.6), the *Simpson's Paradox* occurs in the nonlinear case, for $n > 1$, as well.

• *The case of more than 4 solitons.* Suppose that a contingency table consists of factors A with two levels, factor B with other 2 levels and a third factor C with $l \geq 2$-levels. This case corresponds, in the framework of the *Quantum Harmonic Oscillator*, to considering an initial datum of the form

$$\psi(0,x) = \Sigma_{i=1}^{4l}\psi_i(0,x),$$

with each $\psi_i(0,x)$ of the form:

$$\psi_i(t,x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{i[x\cdot v_i(t)+\gamma_i(t)+\frac{n\omega t}{2}]}e^{-\frac{m\omega}{2\hbar}|x-x_i(t)|^2},$$

where the parameters $x_i(t)$, $v_i(t)$ and $\gamma_i(t)$ are defined as above (see Subsection 3.2.2). In the case of contingency tables, we say that a *Simpson's reversal* occurs when

$$p_{4i+1}p_{4i+4} \geq p_{4i+2}p_{4i+3} \quad \text{for any } i = 1, \dots l,$$

but

$$\left(\Sigma_{i=0}^{l-1}p_{4i+1}\right)\left(\Sigma_{i=0}^{l-1}p_{4i+4}\right) \leq \left(\Sigma_{i=0}^{l-1}p_{4i+2}\right)\left(\Sigma_{i=0}^{l-1}p_{4i+3}\right)$$

or the same occurs, but with all the inequality reversed. In the context of the *Quantum Harmonic Oscillator* with an initial datum of the form $\psi(0,x) = \Sigma_{i=1}^{4l}\psi_i(0,x)$, we say that the *Simpson's Paradox* occurs when

$$Cov\left(\psi_{4i+1}(t,x), \psi_{4i+4}(t,x)\right) > Cov\left(\psi_{4i+2}(t,x), \psi_{4i+3}(t,x)\right) \quad \text{for any } i = 1, \dots l,$$

but

$$Cov\left(\Sigma_{i=1}^{l}\psi_{4i+1}(t,x), \Sigma_{i=1}^{l}\psi_{4i+4}(t,x)\right) < Cov\left(\Sigma_{i=1}^{l}\psi_{4i+2}(t,x), \Sigma_{i=1}^{l}\psi_{4i+3}(t,x)\right).$$

The proof of the occurrence of the *Simpson's Paradox*, in the case of $l > 1$, follows the same lines of the case $l = 1$, with just very small adaptations. We just want to mention that, with the increase of l, the paradox becomes rarer and rarer. See the next paragraph for more details.

• *Probability of the occurrence of the Simpson's Paradox.* An important question is: "How likely is Simpson's Paradox?". It is in fact interesting to quantify, in some way, the chances that one has to run into the paradox. In the case of contingency tables, Pavlides and Perlman address the problem [94] and, among the other things, they prove the following. Suppose that a contingency table consists of a factor A with two levels, a factor B with other 2 levels and a third factor C with $l \geq 2$-levels. Then, the array of cell probabilities $\mathbf{p}$ lies on the Simplex

$$\mathcal{S}_{4l} := \left\{ \mathbf{p} | p_i \geq 0, \ \forall i = 1, \ldots, 4l; \Sigma_{i=1}^{4l} p_i = 1 \right\}.$$

Endow $\mathcal{S}_{4l}$ with the *Dirichlet Distribution on* $\mathcal{S}_{4l}$, denoted by $D_{4l}(\alpha)$ and denote with $\pi_l(\alpha)$ the probability of having the *Simpson's Paradox* under $D_{4l}(\alpha)$. Pavlides and Perlman proved in [94] that $\pi_2(1) = \frac{1}{60}$ and conjectured that for every $\alpha > 0$, there exists $h(\alpha) > 0$ such that

$$\pi_l(\alpha) \simeq \pi_2(\alpha) \times e^{-h(\alpha)\left(\frac{l}{2}-1\right)}, \quad l = 2, 3, \ldots.$$

A similar question can be asked in the case of the *Quantum Harmonic Oscillator* and *Nonlinear Schrödinger equation.* In our constructions, we aimed just at finding one single choice of any parameter which gives the *Simpson's Reversal* and we did it mainly with a perturbative method. But how large is (and in which sense it is large) the set of parameters which gives the *Simpson's Reversal*? One, first, has to construct a reasonable probability distribution on the space of parameters (and, maybe later, in a more advanced way, on the space of solutions) and then compute. One could use something similar to what Pavlides and Perlman presented, but it does not seem reasonable to not use the particular features of the Schrödinger equation. The stationary case seem much simpler than the non-stationary one. It is also interesting to understand if the probability of getting the *Simpson's Paradox* changes varying the spatial dimension n. From our construction, it is not clear if to conjecture that the probability increases or decreases with n, since, for our purposes, it is enough to reduce the problem to the one dimensional case. From one side, in dimension $n \geq 2$, there are more geometrical shapes which can give the paradox, like triangles or rectangles. But, in the same way, there are more sets of parameters which do not give the *Simpson's Reversal.* We think that, since the measure of amalgamation of two solitons increases with the dimension n, the *Simpson's Paradox* should be rarer and rarer. Even if we reported some preliminary results, the problem of the likelihood of the *Simpson's Paradox* remains an open problem.

• *Connection with Game Theory.* The *Prisoner's Dilemma* is an example, in game theory, that shows why two players might decide to not cooperate, even if it appears that it is more convenient for both of them. Suppose we are playing a game with two players A and B. They can choose to either "Cooperate" or "Not-Cooperate". Accordingly to their choices, they get certain payoffs. If both players A and B decide to cooperate, they both receive a reward, p_1. If B chooses to not cooperate while A cooperates, then B receives a payoff p_2, while A receives a payoff p_3. Similarly, if the roles of A and B are reversed. If both players A and B decide to not cooperate, their payoffs are going to be p_4. To get the *Prisoner's Dilemma*, the following must hold on the payoffs:

$$p_3 > p_1 > p_4 > p_2.$$

This condition is the analogue to the original requirement on the *Simpson's Paradox* of having angles in decreasing order. Here $p_3 > p_1$ and $p_4 > p_2$ imply that it is better to not cooperate for both A and B (condition $\theta_3 > \theta_1$ and $\theta_4 > \theta_2$) both given the fact that the other player does or does not cooperate. This situation in game theory is called *Nash Equilibrium*. The dilemma then is that the Nash Equilibrium might not be the global optimal situation, similarly as in the case of the *Simpson's Paradox*, where the *Simpson's Reversal* can occur $\theta_{3,4} < \theta_{1,2}$. We have discussed this in more detail in Section 3.1.

• *Test for Linearity.* Suppose that during a physical experiment, one observes the present of solitons and want to understand if the phenomenon is linear on non-linear. The fact that in the linear case the *Simpson's paradox* is persistent and in the nonlinear it is not, could be use to test linearity against non-linearity. If during the experiment, the paradox appears for almost all the times t (or almost never) then one can think that the solitons are present due to a linear interaction and a presence of an external trapping potential. Instead, if the paradox appears and disappears in a relevant amount of instants t, then the phenomenon cannot be modelled in a linear way and one has to use a nonlinear one. We have not pursued this direction yet.

• *Way to build a continuous of contingency tables without the Paradox.* In the construction of contingency tables, it is very important to try to avoid the Paradox. Our theorems, in some sense, give a way to construct such tables. Theorem 3.2.1 and Theorem 3.2.5 basically say that, with the proper choice of the initial contingency table, one can construct a continuous of contingency tables indexed by t, using the *Quantum Harmonic Oscillator Flow* and that this continuous of contingency tables is actually stable in the sense that, if one varies the initial entries of a small quantity, one goes on avoiding the paradox for any time t and so for all the family of contingency tables.

• *Other Equations.* It would be interesting to verify if the *Simpson's Paradox* occurs also for other PDEs or ODEs. Our construction often relies on the possibility of having a 2-dimensional co-domain for the solutions, but we do not exclude that, with a different *Measure of Amalgamation* to which you can adapt a definition of *Simpson's Paradox*, one can still obtain the *Simpson's Paradox* for equation whose co-domain is, for example, $\mathbb{R}$. We expect that a similar phenomenon could appear, for example, for the *Dirac Equation.*

3.3 The Simpson's Paradox in Big Data and the Consensus Monte Carlo Algorithm

We go back to the problem of distributing a big amount of data to different machines and approach based on the Consensus Monte Carlo algorithm to address it.

We will describe how the algorithm works as explained in [105] and add that the *Simpson's Paradox* appears as a real problem in the "consensus" estimate.

3.3.1 Introduction and Motivation

The Consensus Monte Carlo algorithm is a method of performing approximate Monte Carlo simulation from a Bayesian posterior distribution based on very large data sets. The strategy employed by the method is to divide the data sets into multiple processors.

The possible approaches to attack the problem are mainly multi-core and multi-machine computing. The first can be very effective but does not alleviate bottlenecks and requires non-trivial coding. The second does eliminate bottlenecks, but requires high computational cost in order to communicate between different machines. The two approaches are complementary.

The Consensus Monte Carlo algorithm proposes to divide big data sets across several different machines and sample independently from the posterior of each machine given its own share of data.

The authors of [105] mentioned that a problem with the algorithm is the possible small-sample bias due to the fact that the full data set is divided into several much smaller samples. Jackknife can be used to work on this issue.

In the following, we will add that a further problem is related to the possibility of the *Simpson's Paradox* appearing at different levels of the division of the data into different machines.

In Subsection 3.3.2, we describe the Consensus Monte Carlo (CMC) algorithm, while in Subsection 3.3.3, we highlight the possibility of the appearance of the *Simpson's Paradox* as a problem in the consensus estimate.

3.3.2 The Algorithm

Following [105], we have that the CMC algorithm works in the following way. The complete data set is divided into "shards"; each shard is sent to a machine which performs a Monte Carlo simulation from the posterior distribution given its part of the data; at this point, the estimate of each machine is sent back to another machine which is responsible for the average of the estimate and so producing the "consensus" estimate.

More precisely, suppose that $\mathbf{y}$ represents the full data set, $\mathbf{y}_s$ the data set given to the machine s and θ being the vector of model parameters. Assuming that the sampling procedure is done independently by each machine, but allowing dependence in each machine, we get:

$$p(\theta|\mathbf{y}) \propto \Pi_{s=1}^{S} p(\mathbf{y}_s|\theta) p(\theta)^{1/S}.$$

Here $p(\theta)$ represents the prior information on θ, while $p(\mathbf{y}_s|\theta)$ is the posterior produced by machine s.

Each machine can generate $\theta_{s1}, \ldots, \theta_{sG}$ with $s = 1, \ldots, S$ draws from $p(\theta|\mathbf{y}_s)$. At this point, a machine averages the draws received from each of the s machines with the appropriate weights W_s and obtain the consensus posterior for draw g.

$$\theta_g = \frac{\sum_{s=1}^{S} W_s \theta_{sg}}{\sum_{s=1}^{S} W_s}.$$

The common choices of the weights is the sample variance, but other possible estimates of the variance of the sampling distribution are available.

3.3.3 The Simpson's Paradox for the Consensus Monte Carlo Algorithm

Let us consider as an example the data set relative to the relative to Berkeley Gender Bias in Graduate School Admission described in Subsection 3.1.3.

Suppose that $S = 6$ and each s represents one of the departments A, B, C, D, E, F. Consider the two subgroups $G = Male = 0$ and $G = Female = 0$ and build a logistic regression model for the whole dataset and for each department, having the response variable $Y = 1$ representing the admission, while $Y = 0$ the not admission. We want to give an estimate of the coefficient θ, representing the correlation between Y and the gender difference.

As explained in Subsection 3.1.3, on the whole data set the estimate for θ would be positive in favour of males, while mostly negative in each single department. The apparent contradiction was explained with the presence of a confounding variable, namely the fact that female students applied to more competitive departments, with lower admission rates.

As seen, this issue comes when the data are somehow limited with respect to the current quantity of information available. The possibility of the presence of several confounding variables in big data sets is real and must be taken into consideration in algorithms like the Consensus Monte Carlo where estimates from different machines are combined together. And the probability of incurring in the *Simpson's Paradox* is definitely not negligible (see again [94]).

3.4 Concluding Remarks

In this chapter, we have studied the *Simpson's Paradox*. The *Simpson's Paradox* is the phenomenon that appears in some datasets, where subgroups with a common trend show the reverse trend when they are aggregated. We noticed the occurrence of the paradox in several different areas in science by giving extended examples. The main new results of this chapter concerned the occurrence of the *Simpson's Paradox* in *Quantum Mechanics*. We proved that the *Simpson's Paradox* occurs for solutions of the *Quantum Harmonic Oscillator* both in the stationary case and in the non-stationary case. We proved that the phenomenon is not isolated and that it appears (asymptotically) in the context of the *Nonlinear Schrödinger Equation* as well. Moreover, we discussed the likelihood of occurrence of the paradox in *Quantum Mechanics* and noticed its relation to the *Bohr radius* which might have important physical consequences. It would be good to check if we can detect this phenomenon experimentally.

Chapter 4

Univariate and Multivariate Distributions with Symmetries

This chapter contains some new results about distributions with symmetries.

First, we discuss a result on symmetric order statistics. We prove that the symmetry of any of the order statistics is equivalent to the symmetry of the underlying distribution. Then, we characterize elliptical distributions through group invariance and give some properties. Finally, we study geometric probability distributions on the torus with applications to molecular biology.

4.1 Distributions invariant under Discrete Symmetries

In this section, we discuss results about some probability distributions with discrete symmetries.

4.1.1 The Reflection Group

Most of the material of this subsection can be found in [124], [62] or in [36] and the references there in.

A reflection group is a discrete group which is generated by a set of reflections of a finite dimensional Euclidean space. More precisely, we have the following definition.

Definition 4.1.1 (The Reflection Group)**.** *Let E be a finite-dimensional Euclidean space.*

- *A finite reflection group is a subgroup of the general linear group of E which is generated by a set of orthogonal reflections across hyperplanes passing through the origin.*
- *An affine reflection group is a discrete subgroup of the affine group of E that is generated by a set of affine reflections of E.*

For what concerns us, we will always work with $(\mathbb{R}, +, \cdot)$ as the underlying field, but generalizations to other fields, leading to complex reflection groups or reflection groups over finite fields are possible.

Reflection groups, appear, for example, as the symmetry groups of regular polytopes of the Euclidean space. Reflection Groups are the "concrete" version/ particular cases of the more "abstract" Coxeter Groups [36].

In one dimension, we have $E = \mathbb{R}$ and the only nontrivial group is the simple reflection. Simple reflections correspond to the Coxeter group, $\mathbf{A}_1$, in bracket notation $[3^{1-1}]$, it is of rank 1 and order 2, with Coxeter-Dynkin diagram • and represents the 1-simplex.

In two dimensions, we have $E = \mathbb{R}^2$ and the finite reflection groups are the dihedral groups, which are generated by reflections in two lines that form an angle of $\frac{2\pi m}{n}$ with $m, n \in \mathbb{N}$ relatively prime.

In three dimensions, symmetry groups of the five Platonic solids are finite reflection groups. Note that the symmetry group of the regular dodecahedron and its dual, the regular icosahedron, are the Coxeter group H_3 [37].

All these discrete groups act as isometry on E.

Discrete isometry groups generated by reflections appear also when E is a more general Riemannian manifolds, for example on the sphere $\mathbb{S}^n$ or on the hyperbolic space $\mathbb{H}^n$.

4.1.2 Symmetry of a Distribution via Symmetry of Order Statistics

In this subsection, we establish the following characterization of symmetric absolutely continuous distributions and symmetric discrete distributions. Suppose $X_1, \ldots, X_n$ is a random sample from a distribution with pdf/pmf $f_X(x)$, and $X_{1:n}, \ldots, X_{n:n}$ are the corresponding order statistics. Then, the order statistics are symmetric with respect to the reflection group A_1, namely $X_{r:n} \stackrel{\mathrm{d}}{=} -X_{n-r+1:n}$ for some $r = 1, \ldots, n$, if and only if the underlying distribution is symmetric with respect to A_1, $f_X(x) = f_X(-x)$. Here, $\stackrel{\mathrm{d}}{=}$ means that the two random variables have the same distribution. In the discrete case, we assume the support to be finite.

4.1.2.1 Introduction and Motivation

Suppose $X_1, \ldots, X_n$ are n iid random variables. The corresponding order statistics are the X_i's arranged in non-decreasing order, denoted by $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$. A broad literature has been developed on the study of order statistics and we refer to [2] and [38] for an extended set of references. Here, we concentrate on the problem of characterizing probability distributions through some properties of order statistics.

It is clear that the knowledge of the distribution of $X_{1:1}$ completely determines the distributions $X_{r:n}$ for every r such that $1 \leq r \leq n$ and every $n \in \mathbb{N}$. It also completely determines the marginal and joint distributions of various linear combinations of order statistics. Also, the knowledge of the distribution of $X_{n:n}$ determines the distribution of the X_i's completely. This is true for the following simple reason. To fix the ideas, think about the absolutely continuous case. Since the cdf of $X_{n,n}$ is $F_{n:n}(x) = [F(x)]^n$ for every x and since F is a positive real-valued function, then $F(x) = (F_{n:n}(x))^{\frac{1}{n}}$. The intermediate order statistics completely determine the distribution $F(x)$, In fact,

$$F_{r:n}(x) = Pr(X_{r:n} \leq x) = I_{F(x)}(r, n-r+1),$$

where $I_{1-p}(r, n-r+1) = r\frac{n!}{r!(n-r)!}\int_0^{1-p} t^{r-1}(1-t)^{n-r}dt$ and the inverse function readily gives you $F(x)$.

It is well known that order statistics possess some more interesting distributional properties if the population distribution is symmetric, say about 0. In this case, by using the facts that $f(-x) = f(x)$ and so that $F(-x) = 1 - F(x)$, we have that $X_{r:n} \stackrel{\mathrm{d}}{=} -X_{n-r+1:n}$ for all $r = 1, \ldots, n$ (see, for example, [2]) and similarly for the joint distributions of any sets of order statistics.

What we want to investigate here is the converse of this assertion, namely, if knowing that $X_{r:n} \stackrel{\mathrm{d}}{=} -X_{n-r+1:n}$ for some $r = 1, \ldots, n$ forces the original pdf to be symmetric. i.e. $f_X(x) = f_X(-x)$. In this regard, we have the following result.

Theorem 4.1.2. *Suppose $X_1, \ldots, X_n$ is a random sample from a distribution with pdf (or pmf) $f_X(x)$, and $X_{1:n}, \ldots, X_{n:n}$ are the corresponding order statistics. Then, $X_{r:n} \stackrel{d}{=} -X_{n-r+1:n}$ for some $r = 1, \ldots, n$ if and only if $f_X(x) = f_X(-x)$. Here $\stackrel{d}{=}$ means that the two random variables have the same distribution. In the discrete case, we assume the support to be finite.*

The proof proceeds in two steps. First, we prove that S, the support of $f(x)$, must be symmetric; then, we prove that $f(x) = f(-x)$ on S. The proof follows slightly different lines in the discrete and absolutely continuous case, and so we divided it in the two cases (see Section 4.1.2.3).

The rest of this subsection is organized as follows. In Subsubsection 4.1.2.2, we explain the notation and list some preliminary results about absolutely continuous and discrete order statistics. In Subsubsection 4.1.2.3, we present the complete proof of Theorem 4.1.2.

4.1.2.2 Notation and Preliminaries

In this subsubsection, we list some preliminary results, concerning absolutely continuous and discrete order statistics.

The form of the pdf of $X_{r:n}$ when the population is absolutely continuous is given by the following theorem.

Theorem 4.1.3. *Let $X_{1:n}, \ldots, X_{n:n}$ denote the order statistics of a random sample*

$$X_1, \ldots, X_n$$

from an absolutely continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then, the pdf of $X_{r:n}$ is

$$f_{X_{r:n}}(x) = \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)}[F_X(x)]^{r-1}[1 - F_X(x)]^{n-r} f_X(x). \tag{4.1}$$

The form of the pmf of $X_{r:n}$ when the population is discrete is given by the following theorem.

Theorem 4.1.4. *Let $X_{1:n}, \ldots, X_{n:n}$ denote the order statistics of a random sample*

$$X_1, \ldots, X_n$$

from a discrete population with cdf $F_X(x)$ *and pmf* $f_X(x)$*. Then, the pmf of* $X_{r:n}$ *is*

$$P(X_{r:n} = x) = \sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!}$$
$$\left((1 - F(x))^j (F(x))^{n-j} - (1 - F(x) + f(x))^j (F(x) - f(x))^{n-j}\right).$$

For more details on the distribution theory of order statistics, we refer to [2] and [38].

4.1.2.3 Proof of Theorem 4.1.2

In this subsubsection, we present a complete proof of Theorem 4.1.2. One direction is trivial, as already explained in the introduction.

Absolutely Continuous Case

Now, we consider the case in which the distribution is absolutely continuous. We start by proving that the support must be symmetric.

Lemma 4.1.5 (Symmetric Support)**.** *Suppose* $X_1, \ldots, X_n$ *is a random sample from an absolutely continuous density* $f_X(x)$*. If* $X_{r:n} \overset{d}{=} -X_{n-r+1:n}$*, then* $S := \{x : f_X(x) > 0\}$*, the support of* $f_X(x)$*, is symmetric.*

Proof. Suppose $X_{r:n} \overset{d}{=} -X_{n-r+1:n}$. Then, $f_{X_{r:n}}(x) = f_{X_{n-r+1:n}}(-x)$ and so

$$\frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)}[F_X(x)]^{r-1}[1 - F_X(x)]^{n-r} f_X(x)$$
$$= \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)}[F_X(-x)]^{n-r}[1 - F_X(-x)]^{r-1} f_X(-x).$$

If $x \in S^c$, then $f_X(x) = 0$. Consequently, by the condition $f_{X_{r:n}}(x) = f_{X_{n-r+1:n}}(-x)$, we get

$$[F_X(-x)]^{n-r}[1 - F_X(-x)]^{r-1} f_X(-x) = 0.$$

On the left hand side, we have the product of three terms. Therefore, in order for the left hand side to be zero, it is enough that at least one of the factors is zero. Then, we have three cases:

- $f(-x) = 0$ and so $-x \in S^c$;
- $F(-x) = 0$ which implies $f(-x) = 0$ and so $-x \in S^c$;
- $F(-x) = 1$ which implies $f(-x) = 0$ and so $-x \in S^c$.

In all the three cases, we get that if we suppose that $x \in S^c$, then $-x \in S^c$. If we repeat the same steps assuming now that $-x \in S^c$, we will get $x \in S^c$. Therefore, $-x \in S^c$ if and only if $x \in S^c$ and so $x \in S$ if and only if $-x \in S$. This means that the support S is symmetric, as required. $\square$

Now, we can work on S the support of $f(x)$.

By Lemma 4.1.5, we can concentrate on $x \in S$. By Theorem 4.1.3, and by assumption

$$\begin{aligned}&\frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)}[F_X(x)]^{r-1}[1-F_X(x)]^{n-r}f_X(x)\\&=\frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)}[F_X(-x)]^{n-r}[1-F_X(-x)]^{r-1}f_X(-x).\end{aligned}$$

This implies

$$[F_X(x)]^{r-1}[1-F_X(x)]^{n-r}f_X(x)=[F_X(-x)]^{n-r}[1-F_X(-x)]^{r-1}f_X(-x).$$

Since $x \in S$, we can divide term by term by $[F_X(-x)]^{n-r}[1-F_X(-x)]^{r-1}f_X(x)$ and get:

$$\left[\frac{F_X(x)}{1-F_X(-x)}\right]^{r-1}\left[\frac{1-F_X(x)}{F_X(-x)}\right]^{n-r}=\frac{f_X(-x)}{f_X(x)}.$$

By Lemma 4.1.5 and changing $x \mapsto -x$ in this identity, we get the symmetric relation

$$\left[\frac{F_X(-x)}{1-F_X(x)}\right]^{r-1}\left[\frac{1-F_X(-x)}{F_X(x)}\right]^{n-r}=\frac{f_X(x)}{f_X(-x)}.$$

If we multiply term by term, we get

$$\left[\frac{F_X(x)}{1-F_X(-x)}\right]^{-n+2r-1}\left[\frac{1-F_X(x)}{F_X(-x)}\right]^{n-2r+1}=1$$

and so

$$\left[\frac{1-F_X(x)}{F_X(-x)}\right]^{n-2r+1}=\left[\frac{F_X(x)}{1-F_X(-x)}\right]^{n-2r+1}.$$

Since for $x \in S$, we have $0 < F(x) < 1$, we get:

$$\frac{1-F_X(x)}{F_X(-x)}=\frac{F_X(x)}{1-F_X(-x)}.$$

Multiplying term by term, we get

$$[1 - F_X(x)][1 - F_X(-x)] = [F_X(-x)][F_X(x)],$$

which simplifying becomes

$$F_X(-x) = 1 - F_X(x),$$

which is equivalent to $f_X(x) = f_X(-x)$. This concludes the proof in the absolutely continuous case.

Discrete Case

Now, we consider the case in which the distribution is discrete.

We start by proving that the support must be symmetric.

Lemma 4.1.6 (Symmetric Support). *Suppose $X_1, \ldots, X_n$ is a random sample from a discrete density $f_X(x)$. If $X_{r:n} \overset{d}{=} -X_{n-r+1:n}$, then $S := \{x : f_X(x) > 0\}$, the support of $f_X(x)$, is symmetric.*

Proof. Suppose there exists x such that $Pr(X = x) = p_x > 0$, while $Pr(X = -x) = 0$. This already implies $x \neq 0$. Since $|S| < +\infty$, then there exists a well defined x_m such that

$$x_m = arg \max\{x : Pr(X = x) > 0, \ \ but \ Pr(X = -x) = 0\}.$$

For simplicity we call $-x = x_1$ and $x = x_m$, with $Pr(X = x_m) = p_m$.

By hypothesis $P(X_{r:n} = x) = P(X_{n-r+1} = -x)$. By Theorem 4.1.4 and by assumption, we get

$$\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} \left((1 - F(x))^j (F(x))^{n-j} - (1 - F(x) + f(x))^j (F(x) - f(x))^{n-j}\right)$$

$$= \sum_{j=0}^{r-1} \frac{n!}{j!(n-j)!}$$

$$\times \left((1 - F(-x))^j (F(-x))^{n-j} - (1 - F(-x) + f(-x))^j (F(-x) - f(-x))^{n-j}\right).$$

By the binomial identity $(a+b)^n = \sum_{j=0}^{n} \frac{n!}{j!(n-j)!} a^j b^{n-j}$ with $a = 1-F(-x)$ and $b = F(-x)$, and, respectively, with $a = 1 - F(-x) + f(-x)$ and $b = F(-x) - f(-x)$, we get

$$\sum_{j=0}^{r-1} \frac{n!}{j!(n-j)!} \left((1-F(-x))^j (F(-x))^{n-j} - (1-F(-x)+f(-x))^j (F(-x)-f(-x))^{n-j}\right)$$

$$= -\sum_{j=r}^{n} \frac{n!}{j!(n-j)!}$$

$$\times \left((1-F(-x))^j (F(-x))^{n-j} - (1-F(-x)+f(-x))^j (F(-x)-f(-x))^{n-j}\right)$$

$$= -\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!}$$

$$\times \left((1-F(-x))^{n-j} (F(-x))^{j} - (1-F(-x)+f(-x))^{n-j} (F(-x)-f(-x))^{j}\right),$$

by changing index $j \mapsto n-j$. Therefore, we have the identity

$$\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} \left((1-F(x))^j (F(x))^{n-j} - (1-F(x)+f(x))^j (F(x)-f(x))^{n-j}\right)$$

$$= -\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!}$$

$$\times \left((1-F(-x))^{n-j} (F(-x))^{j} - (1-F(-x)+f(-x))^{n-j} (F(-x)-f(-x))^{j}\right),$$

which must be true for every $x \in S$. For $x = x_m$, we get

$$\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} \left(0^j 1^{n-j} - p_m^j (1-p_m)^{n-j}\right) = -\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} \left(p_1^j (1-p_1)^{n-j} - (1)^{n-j} 0^j\right),$$

and so

$$\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} p_m^j (1-p_m)^{n-j}$$

$$= \sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} p_1^j (1-p_1)^{n-j} = 0.$$

This implies that each term of the sum is zero, since already non-negative, which means $p_m = 0$ or $p_m = 1$.

If $p_m = 1$, we have that $Pr(X = x_m) = 1$ and so $Pr(X_{r:n} = x_m) = 1$ for every $r = 1, \ldots, n$.

Since, by assumption, $P(X_{r:n} = x) = P(X_{n-r+1} = -x)$, then $x = -x = x_m = 0$, which is excluded, for the same argument at the beginning of the proof. Therefore $p_m = 0$.

The same argument can be repeated on the set $S \setminus \{x_1, x_m\}$ and then iteratively. This concludes the proof of the lemma. $\square$

Now, we can work on S the support of $f(x)$. By Lemma 4.1.6, we can concentrate on $x \in S$. By hypothesis $P(X_{r:n} = x) = P(X_{n-r+1} = -x)$.

By the same computations in the proof of Lemma 4.1.6, we are reduced to the identity

$$\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} \left((1-F(x))^j (F(x))^{n-j} - (1-F(x)+f(x))^j (F(x)-f(x))^{n-j}\right)$$

$$= -\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!}$$

$$\times \left((1-F(-x))^{n-j}(F(-x))^j - (1-F(-x)+f(-x))^{n-j}(F(-x)-f(-x))^j\right),$$

which must be true for every $x \in S$.

Since $|S| < +\infty$, we have $S = \{x_1, \ldots, x_m\}$ for some $m \in \mathbb{N}$, and so that $\sum_{i=1}^m p_i = 1$, where $p_i = P(X = x_i)$ for $i = 1, \ldots m$ and $x_i = -x_{m-i+1}$ for $i = 1, \ldots m$ by Lemma 4.1.6.

Consider $x = x_m$, we get

$$\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} \left(0^j 1^{n-j} - p_m^j (1-p_m)^{n-j}\right) = -\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} \left(p_1^j (1-p_1)^{n-j} - (1)^{n-j} 0^j\right),$$

and so

$$\sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} p_m^j (1-p_m)^{n-j} = \sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} p_1^j (1-p_1)^{n-j}.$$

Consider the function

$$g(p) := \sum_{j=0}^{n-r} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j}.$$

Using well known results about the regularized incomplete Beta function, we get

$$g(p) = r\frac{n!}{(n-r)!r!}\int_0^{1-p} t^{r-1}(1-t)^{n-r}dt.$$

This function is differentiable in p for $p \in [0,1]$ and so

$$g'(p) = -r\frac{n!}{(n-r)!r!}(1-p)^{r-1}p^{n-r} \leq 0.$$

Note that $g'(p) = 0$ if and only if $p = 0$ or $p = 1$. Note that $p = 0$ if and only if we are outside the support (but we are on the support) and $p = 1$ in the degenerate case of $|S| = m = 1$, when the theorem is trivially true, since $S = \{0\}$ and $P(X = 0) = 1$.

Evaluating $g(p)$ at p_1 and p_m, we get

$$g(p_1) = g(p_m),$$

which, by monotonicity and so invertibility of $g(p)$ for $0 < p < 1$, implies

$$p_1 = p_m.$$

Now, we proceed by induction. Suppose $p_j = p_{m-j+1}$ for every $j = 1, \dots, k-1$. Consider x_k, p_k and x_{m-k+1}, p_{m-k+1}. We get:

$$\sum_{j=0}^{n-r}\frac{n!}{j!(n-j)!}\left((1-F(x_k))^j(F(x_k))^{n-j} - (1-F(x_k)+f(x_k))^j(F(x_k)-f(x_k))^{n-j}\right)$$

$$= -\sum_{j=0}^{n-r}\frac{n!}{j!(n-j)!}\left((1-F(x_{m-k+1}))^{n-j}(F(x_{m-k+1}))^j\right)$$

$$+\sum_{j=0}^{n-r}\frac{n!}{j!(n-j)!}\left((1-F(x_{m-k+1})+f(x_{m-k+1}))^{n-j}(F(x_{m-k+1})-f(x_{m-k+1}))^j\right),$$

which by definition of cdf, becomes

$$\sum_{j=0}^{n-r}\frac{n!}{j!(n-j)!}(p_{k+1}+\cdots+p_m)^j(p_1+\cdots+p_k)^{n-j}$$

$$-\sum_{j=0}^{n-r}\frac{n!}{j!(n-j)!}(p_k+\cdots+p_m)^j(p_1+\cdots+p_{k-1})^{n-j}$$

$$=\sum_{j=0}^{n-r}\frac{n!}{j!(n-j)!}(p_1+\cdots+p_{m-k})^j(p_{m-k+1}+\cdots+p_m)^{n-j}$$

$$-\sum_{j=0}^{n-r}\frac{n!}{j!(n-j)!}(p_{m-k+2}+\cdots+p_m)^{n-j}(p_1+\cdots+p_{m-k+1})^j.$$

By induction hypothesis

$$p_1+\cdots+p_{k-1}=p_{m-k+2}+\cdots+p_m,$$

which implies

$$p_k+\cdots+p_m=1-(p_1+\cdots+p_{k-1})=1-(p_{m-k+2}+\cdots+p_m)=p_1+\cdots+p_{m-k+1}.$$

Therefore, the negative terms cancel out.

We are reduced to

$$\sum_{j=0}^{n-r}\frac{n!}{j!(n-j)!}(p_{k+1}+\cdots+p_m)^j(p_1+\cdots+p_k)^{n-j}$$

$$=\sum_{j=0}^{n-r}\frac{n!}{j!(n-j)!}(p_1+\cdots+p_{m-k})^j(p_{m-k+1}+\cdots+p_m)^{n-j}.$$

We can use again the invertibility of the function $g(p)$ to get

$$p_1+\cdots+p_k=p_{m-k+1}+\cdots+p_m,$$

which, by induction hypothesis, implies

$$p_k=p_{m-k+1}.$$

Note that this concludes the proof in the discrete case, since in the case $|S|$ is even, the induction stops at $k=m/2$, while if $|S|$ is odd the induction stops at $k=(m-1)/2$, with the only constraint of $x_{(m-1)/2+1}=0$ and no extra constraint on the spare $p_{(m-1)/2+1}$.

This concludes the proof of the theorem in both cases.

4.2 Distributions invariant under Continuous Symmetries

Continuous symmetries are usually studied by means of topological groups or Lie groups. Topological groups are groups with a topology with respect to which the group operations are continuous. Lie groups are groups that have the structure of a differentiable manifold, with the group operations are compatible with the differentiable structure. Here are the precise definitions. See for example [42].

Definition 4.2.1. *A topological group $(G, \dot{,} \tau)$ is a topological space with topology τ and is also a group with respect to the operation $\cdot$ such that the group operations:*

$$(G \times G, \tau \times \tau) \mapsto (G, \tau) : (g_1, g_2) \mapsto g_1 \cdot g_2$$

and

$$(G, \tau) \mapsto (G, \tau) : g \mapsto g^{-1}$$

are continuous with respect to the proper topology. Here $g_1, g_2, g \in G$.

Definition 4.2.2. *A real Lie group $(G, \dot{,} \mathcal{A})$ is real differentiable manifold with differentiable structure $\mathcal{A}$, and a group with respect to the operation $\cdot$ such that the group operations:*

$$(G \times G, \mathcal{A} \times \mathcal{A}) \mapsto (G, \mathcal{A}) : (g_1, g_2) \mapsto g_1 \cdot g_2$$

and

$$(G, \mathcal{A}) \mapsto (G, \mathcal{A}) : g \mapsto g^{-1}$$

are smooth maps with respect to the proper differentiable structures. Here $g_1, g_2, g \in G$.

One of the most important of these groups is the orthogonal group (see Definition 4.3.1 below).

A lot of very important distributions are invariant under the action of this group, like spherical distributions and the Standard Multivariate Normal in particular.

In the next section, we will study the characterization of elliptical distributions, a more flexible version of spherical distributions, through group invariance.

4.3 A characterization of elliptical distributions through group invariance

In this section, we present a characterization of elliptical distributions through group invariance. We complete the analysis characterizing also marginals and conditional distributions of elliptical distributions as invariant under subgroups of the original group. We continue with an appendix on further properties like the polar decomposition and the existence of the Haar Measure. Finally, we cast our discussion in the context of the metric ground form.

4.3.1 Introduction and Motivation

Spherical and elliptical distributions are flexible extensions of the multivariate normal distribution that keep the (probably) most important feature of Gaussians, namely the ellipticity of the probability contours. Spherical and elliptical distributions have been studied in extended detail and we refer the interested reader to the works of Fang-Kotz-Ng [51], Eaton [47], Cambanis-Huang-Simons [23], Diaconis [39], Fang-Anderson [50], Gomez-Gomez-Marin [58] and the references therein.

A natural definition of spherically symmetric distribution is through their invariance with respect to the group orthogonal transformations.

Definition 4.3.1. *An $p \times 1$ random vector $\mathbf{X}$ is said to be* spherically symmetric *distributed if and only if*

$$\mathbf{X} \overset{d}{=} H\mathbf{X}$$

for every $H \in \mathcal{O}(p)$, where $\mathcal{O}(p)$ is the orthogonal group *(see Section 4.3.4 for the definition of $\mathcal{O}(p)$).*

Elliptical distributions instead, are usually defined by means of linear transformations of spherical distributions [47] or assuming the existence of a density of a particular form [90] or through the particular form of their characteristic function [47], [90].

In this subsection, we give an equivalent but group theoretical definition of elliptical distributions, as it is commonly done for spherical distributions. We restrict our attention to elliptical distributions, centred in the origin and with symmetry axes parallel to the coordinate axes in order to keep the parallel with orthogonal transformations. Technically, this means that we are working in the quotient space with respect to rigid motions. In fact, the general case of non-zero mean and non-trivial correlations can be recovered by rigid transformations of the Euclidean space. This is consistent with the common choice in the definition of spherical distributions, to be spherical with respect to the origin. We introduce what we call *Stretched Orthogonal Matrices*.

Definition 4.3.2. *Suppose that we have a sequence of real numbers $\lambda_i \in \mathbb{R}$ for every $i = 1, \dots, p$ such that $\lambda_1^2 \times \dots \times \lambda_p^2 = 1$ and that $\Lambda := diag(\lambda_1^2, \dots, \lambda_p^2)$. Then, we define the set of* Λ-Stretched Orthogonal Matrices $O_s(p;\Lambda)$ *by*

$$O_s(p;\Lambda) := \left\{ K \in Mat_{p\times p} : K^T \Lambda K = \Lambda \right\}.$$

The set of *Stretched Orthogonal Matrices* possesses a group structure and is diffeomorphic to $O(p)$, but it is not a subgroup of $O(p)$ (apart in the trivial case $\Lambda = Id_{p\times p}$). Moreover, $O_s(p;\Lambda)$ possesses slightly different algebraic properties. We refer to Section 4.3.6 for the precise analysis of these properties.

By means of *Stretched Orthogonal Matrices*, we can characterize elliptical distributions by group invariance. We prove the following theorem.

Theorem 4.3.3. *Suppose that $\mathbf{X} = (X_1, \dots, X_p)^T$ is a random vector. The following statements are equivalent:*

- $\mathbf{X} \overset{d}{=} K\mathbf{X}$ *for every $K \in O_s(p;\Lambda)$.*
- *The characteristic function of $\mathbf{X}$ satisfies $\phi_{\Lambda,\mathbf{X}}(\mathbf{t}) = \phi_{\Lambda,\mathbf{X}}(K^T\mathbf{t})$ for some $\phi_\Lambda \in \Phi_p$ and for any $K \in O_s(p;\Lambda)$).*
- *There exists a function $\psi(\cdot)$ of a scalar variable such that $\psi(\mathbf{t}^T\Lambda\mathbf{t}) = \phi_\Lambda(\mathbf{t})$.*
- *If the density exists, it must take the form*

  $$f_{\mathbf{X}}(\mathbf{x}) = c_p det(\Lambda)^{-1/2} e\left(\mathbf{x}^T\Lambda\mathbf{x}\right),$$

  with c_p a normalizing constant dependent just on the dimension p and the function e. Also, the function e can be taken independent of p.

Note that the equivalence of the last three characterizations is well known, while the first one has not been developed explicitly in the literature, as far as we know.

We have the following definition.

Definition 4.3.4. *A random variable $\mathbf{Y}$ which satisfies one of the previous conditions is said to be a* Λ-Elliptical Distribution *and its distribution will be denoted by $E(\Lambda)$.*

Remark 4.3.5. *In the case $\Lambda = Id_{p\times p}$, the random variable $\mathbf{Y}$ is invariant under $\mathcal{O}(p)$ and so it is simply a* Spherical Distribution*.*

Marginals and conditional distributions of elliptical distributions can be characterized by subgroup invariance. We have the following theorem.

Theorem 4.3.6. *Suppose that an $p\times 1$ random vector $\mathbf{Y}$ follows a* Λ-Elliptical Distribution *and is partitioned as $\mathbf{Y} = (\mathbf{Y}_1^T, \mathbf{Y}_2^T)^T$ with $\mathbf{Y}_1$ an $p_1 \times 1$ vector and $\mathbf{Y}_2$ an $p_2 \times 1$ vector ($p_2 := p - p_1$). Then, the following statements are true:*

- $\mathbf{Y}_1 \stackrel{d}{=} K\mathbf{Y}_1$ *for every $K \in O_s(p_1; \Lambda_{p_1})$.*
- *The group $O_s(p_1; \Lambda_{p_1})$ which leaves invariant $\mathbf{Y}_1$ is the subgroup of $O_s(p; \Lambda_p)$, which does not move the remaining components $\mathbf{Y}_2$. Here $\Lambda_k = diag(\lambda_1^2, \ldots, \lambda_k^2)$.*

A similar statement holds for conditional distributions $\mathbf{Y}_1 \mid \mathbf{Y}_2$.

Remark 4.3.7. *Note that the theorem works also if the components of the vector $\mathbf{Y}_1$ are not necessarily in increasing order of the indexes.*

Remark 4.3.8. *This result can be potentially used as a way to test for spherical symmetry in the class of elliptical distributions. We refer to [8], [10] and [67], for more details on testing for spherical symmetry. An interesting treatment of the distributions on the hyperboloid is given in [64].*

The remaining part of this subsection is organized as follows. In Subsubsection 4.3.2, we collect some notation and preliminary results. In particular, we introduce some basic notions in group theory focusing on the group $\mathcal{O}(p)$ of orthogonal matrices and then recall some known results about marginals and conditional distributions of elliptical distributions. In Subsubsection 4.3.6, we discuss the group theoretical properties of *Stretched Orthogonal Matrices* and give their explicit form in the case $p = 2$. In Subsubsection 4.3.7, we give the proofs of our main theorems. We conclude with the Appendix 4.3.8, where we discuss some further properties of *Stretched Orthogonal Matrices*, like the Polar Representation, some Angle-Radius theorems and the construction of the Haar Measure. We also connect our constructions to the metric ground form.

4.3.2 Notations and Preliminaries

In this subsubsection, we fix some notations and discuss some preliminaries. In the following, we use the notation $\Lambda_k := diag(\lambda_1^2, \ldots, \lambda_k^2)$.

4.3.3 Basics on Group Theory and Invariance under Graoup Actions

The section is devoted to a collection of definitions and theorems in *Group Theory*. The fundamental example is the one of *Orthogonal Matrices* $\mathcal{O}(p)$. We focus on what we need for our main theorems and, for more details on group theoretical approaches in statistics, we refer to Fang-Kotz-Ng [51], Eaton [47] and Diaconis [39].

Definition 4.3.9. *The couple* $(G, \cdot)$ *is called* Group*, if the following conditions are satisfied.*

- G *is a set.*
- Closure*:* $\cdot$ *is a binary operation, defined on the set* G*:*

  $$\cdot : G \times G \to G : (g_1, g_2) \mapsto g_1 \cdot g_2,$$

  such that for every $g_1, g_2 \in G$*, there exists* $g_3 \in G$ *with* $g_3 = g_1 \cdot g_2$*.*
- Associativity*: For all* $g_1, g_2, g_3 \in G$ *the following condition holds true:*

  $$g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3.$$

- Identity*: There exists* $e \in G$ *such that for every* $g \in G$*, the following condition holds true:* $e \cdot g = g \cdot e = g$*.*
- Inverse*: For every* $g \in G$*, there exists* h_g *often denoted by* $h_g := g^{-1}$ *such that* $g \cdot h_g = h_g \cdot g = e$*.*

Definition 4.3.10. *Suppose that* $(G, \cdot)$ *is a Group and* $\mathcal{X}$ *is a set,. A* Left Group Action Φ of G on X *is a function*

$$\Phi : G \times \mathcal{X} \to \mathcal{X} : (g, x) \mapsto \Phi(g, x)$$

that satisfies the following two conditions:

- Compatibility*: For every* $g_1, g_2 \in G$ *and every* $x \in \mathcal{X}$*, we have that*

  $$\Phi(g_2, \Phi(g_1, x)) = \Phi(g_2 \cdot g_1, x).$$

- Identity*: For every* $x \in \mathcal{X}$ *and with* e *the Identity Element of* G *we have that:*

  $$\Phi(e, x) = x.$$

Remark 4.3.11. *Note that here, differently from [51], we are distinguishing between Group and Group Action.*

In the following, we will often write g_1g_2 instead of $g_1 \cdot g_2$.

Definition 4.3.12. *Two points $x_1, x_2 \in \mathcal{X}$ are said to be* equivalent *under G, if there exists $g \in G$ such that $x_2 = gx_1$. In such a case, we write $x_1 \sim x_2$ mod G.*

Proposition 4.3.13. *The relation $\sim$ is an* equivalence relation *in G, in the sense that it satisfies the following properties:*

- Reflexivity*: For any $x \in X$, it holds that $x \sim x$ mod G.*
- Symmetry*: For any $x_1, x_2 \in X$, if it holds that $x_1 \sim x_2$ mod G, then it holds that $x_2 \sim x_1$ mod G.*
- Transitivity*: For any $x_1, x_2, x_3 \in X$, the following condition holds: if $x_1 \sim x_2$ mod G and $x_2 \sim x_3$ mod G, then $x_1 \sim x_3$ mod G.*

To enlighten the notation, we will often write $x_1 \sim x_2$ instead of $x_1 \sim x_2 \bmod G$.

Definition 4.3.14. *A function $f : \mathcal{X} \to \mathbb{R}$ is said to be* Invariant under G*, if the following condition holds:*

$$f(gx) = f(x) \text{ for each } x \in \mathcal{X} \text{ and each } g \in G.$$

A function $f : \mathcal{X} \to \mathbb{R}$ is said to be Maximal Invariant under G*, if it is Invariant under G and if the following condition holds:*

$$f(x_1) = f(x_2) \text{ implies } x_1 \sim x_2.$$

Definition 4.3.15. *The* Orbit $O(x)$ of $x \in \mathcal{X}$ under the Group $(G, \cdot)$ *is defined by the following set:*

$$O(x); = \{g \cdot x \mid g \in G\}.$$

By these last two definitions, we can say that a *Maximal Invariant Function* is a function which assumes different values for different orbits.

Proposition 4.3.16. *Consider a set $\mathcal{X}$ and a group $(G, \cdot)$ which acts on X from the left. Assume that $f : \mathcal{X} \to Y_1$ is Maximal Invariant under G. Then a function $h : \mathcal{X} \to Y_2$ is Maximal Invariant under G if and only if there exists $\nu : Y_1 \to Y_2$ such that $h(x) = \nu(f(x))$.*

Proof. See [51] (page 15, Theorem 1.1). □

We define distributions invariant under the action of a group.

Definition 4.3.17. *Let X be a random variable with values in the sample space $\mathcal{X}$ and whose pdf is P_θ with parameter space $\theta \in \Omega$. Let G be a group of transformations from $\mathcal{X}$ into itself. The family of distributions $\{P_\theta; \theta \in \Omega\}$ is said to be* Invariant under G *if every $g \in G, \theta \in \Omega$ determine a unique element $\bar{g}\theta$ in Ω, such that, when $X \sim P_\theta$, then $gX \sim P_{\bar{g}\theta}$, namely:*

$$P_{\bar{g}\theta}(gB) = P_\theta(B)$$

for every $B \subset X$ measurable.

4.3.4 The example of *Orthogonal Matrices*

In this section, we present the group of *Orthogonal Matrices* which naturally appear in the study of spherical distributions.

Let $O(p) := \{H \in Mat_{p\times p} : H^T H = Id_p\}$ be the set of orthogonal matrices. Suppose that $\mathcal{X} = \mathbb{R}^p$ and $G = O(p)$ with the matrix multiplication $\cdot$ as the operation. It is a very well known fact that $(G, \cdot)$ is a group which acts on the left on $\mathcal{X}$.

Proposition 4.3.18. *The function $f : \mathbb{R}^p \to \mathbb{R}$ given by $f(x) = \|x\|^2$ is Maximal Invariant under $(G, \cdot)$.*

Proof. The function f is invariant under $(O(p), \cdot)$, in fact

$$f(Hx) = x^T H^T Hx$$

for all $x \in \mathbb{R}^p$ and for all $H \in O(p)$. On the other hand, if $f(x_1) = f(x_2)$ for some $x_1, x_2 \in \mathbb{R}^p$, then there exists $H \in O(p)$ such that $x_2 = Hx_1$. This is true because two points at the same distance form the origin can be moved one to the other by a rotation. This implies that $x_1 \sim x_2$ and so that f is *Maximal Invariant.* □

Corollary 4.3.19. *Suppose that $x \in \mathbb{S}^{p-1} := \{x \in \mathbb{R}^p : x^T x = 1\}$. Then, for any $H \in O(p)$, we have that $Hx \in \mathbb{S}^{p-1}$, namely the $\mathbb{S}^{p-1}$ is invariant under the action of $O(p)$.*

Proof. It is a simple application of the previous proposition. □

Remark 4.3.20. *Given a group $(G, \cdot)$, there might be several* Maximal Invariant *functions under $(G, \cdot)$. For example, in the case just discussed, any $g(x) = a\|x\|^2$ with $a > 0$ is* Maximal Invariant *under $(O(p), \cdot)$. Also, $g(x) = h(f(x))$ for any $h \in C^1(\mathbb{R})$ with $h' > 0$.*

4.3.5 Marginals and Conditionals of Elliptical Distributions

Here, we collect a couple of theorems from [90] concerning marginals and conditionals of elliptical distributions. We start with marginal distributions.

Theorem 4.3.21. *Suppose that* $\mathbf{X}$ *is elliptically distributed with mean* μ *and variance* V*. Here* μ *are* $p \times 1$ *vectors and* V *is an* $p \times p$ *symmetric positive definite matrix. Suppose that* $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T$ *with* $\mathbf{X}_1$ *a* $p_1 \times 1$ *vector and that* $\mu = (\mu_1^T, \mu_2^T)^T$ *with* μ_1 *a* $p_1 \times 1$ *and also that* $V = [V_{11}, V_{12}; V_{21}, V_{22}]$ *is the corresponding partitioned matrix. Then* $\mathbf{X}_1$ *is elliptically distributed with mean* μ_1 *and variance* V_{11}*.*

We continue with conditional distributions.

Theorem 4.3.22. *Suppose that* $\mathbf{X}$ *is elliptically distributed with mean* μ *and variance* V*. Here* μ *are* $p \times 1$ *vectors and* V *is an* $p \times p$ *symmetric positive definite matrix. Suppose that* $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T$ *with* $\mathbf{X}_1$ *a* $p_1 \times 1$ *vector and that* $\mu = (\mu_1^T, \mu_2^T)^T$ *with* μ_1 *a* $p_1 \times 1$ *and also that* $V = [V_{11}, V_{12}; V_{21}, V_{22}]$ *is the corresponding partitioned matrix. Then* $\mathbf{X}_1$ *is elliptically distributed with mean*

$$E[\mathbf{X}_1 \mid \mathbf{X}_2] = \mu_1 + V_{12} V_{22}^{-1} (\mathbf{X}_2 - \mu_2)$$

and variance

$$Cov[\mathbf{X}_1 \mid \mathbf{X}_2] = g(\mathbf{X}_2) \left(V_{11} - V_{12} V_{22}^{-1} V_{21} \right)$$

for some function g*.*

4.3.6 Stretched Orthogonal Matrices

In this section, we discuss *Stretched Orthogonal Matrices* and prove their algebraic properties.

4.3.6.1 Group Theoretical Properties

In this subsection we discuss the group theoretical properties of *Stretched Orthogonal Matrices*. First of all, we have that $(O_s(p; \Lambda)), \cdot)$ is a group.

Theorem 4.3.23. *The set* $O_s(p; \Lambda)$ *endowed with the matrix multiplication* $\cdot$ *inherited from* $Mat_{p \times p}$ *is a group* $(O_s(p; \Lambda), \cdot)$*.*

Proof. We need to show that $(O_s(p; \Lambda)), \cdot)$ satisfies the properties of Definition 4.3.9.

- *Closure*: Suppose that $K, H \in O_s(p;\Lambda)$. Therefore $K^T \Lambda K = \Lambda$ and $H^T \Lambda H = \Lambda$. We want to verify that
$$(KH)^T \Lambda (KH) = \Lambda.$$
By definition of transpose:
$$(KH)^T \Lambda (KH) = H^T K^T \Lambda K H = H^T \Lambda H = \Lambda,$$
by hypothesis on K and H. So, $(O_s(p;\Lambda), \cdot)$ is closed.

- *Associativity*: This is a consequence of the fact that $O_s(p;\Lambda)$ inherits the operation $\cdot$ from $Mat_{p\times p}$.

- *Identity*: The element $e := Id_{p\times p} \in O_s(p;\Lambda)$ since $Id_{p\times p}^T \Lambda Id_{p\times p} = \Lambda$ and $K Id_{p\times p} = Id_{p\times p} K = K$. Therefore, $e := Id_{p\times p}$ is the Identity Element of $O_s(p;\Lambda), \cdot)$.

- *Inverse*: Since $\lambda_1^2 \times \dots \lambda_p^2 = 1$, then any $K \in O_s(p;\Lambda)$ is invertible. In fact:
$$det(K)^2 = det(K^T)det(\Lambda)det(K) = det(K^T \Lambda K) = det(\Lambda) = 1$$
and so $det(K) \neq 0$. This implies that there exists an inverse $K^{-1} \in Mat_{p\times p}$. Now, we have to prove that $K^{-1} \in O_s(p;\Lambda)$. Since $K \in O_s(p;\Lambda)$, then
$$K^T \Lambda K = \Lambda.$$
Now, multiply by $(K^T)^{-1}$ (which exists since $det(K^T) = det(K)$) on the left and by K^{-1} on the right to get:
$$\Lambda = (K^T)^{-1} \Lambda K^{-1} = (K^{-1})^T \Lambda K^{-1},$$

which is true since $(K^{-1})^T = (K^T)^{-1}$ (just transpose the identity $KK^{-1} = Id_{p\times p}$). But then:
$$(K^{-1})^T \Lambda K^{-1} = \Lambda$$
and so $K^{-1} \in O_s(p;\Lambda)$. □

The following theorem proves that the groups $O(p)$ and $O_s(p;\Lambda)$ are distinct and satisfy some rigidity properties.

Theorem 4.3.24. *It holds that $O_s(p;\Lambda) = O(p)$ if and only if $\lambda_1^2 = \dots = \lambda_p^2$. Moreover,*

- *if* $O_s(p;\Lambda) \neq O(p)$*, then there exists* k *such that* $(O_s(p;\Lambda))_{(i_1,\dots,i_k)\times(i_1,\dots,i_k)} \cap O(k) = \{Id_{k\times k}\}$*;*

- $(O_s(p;\Lambda))_{(i_1,\dots,i_k)\times(i_1,\dots,i_j)} \cap O(j) = O(j) = O_s(j, \lambda_{i_1}, \dots, \lambda_{i_j})$ *for every* $i_1, \dots, i_j$ *such that* $\lambda_{i_1} = \dots = \lambda_{i_j}$.

Remark 4.3.25. *Geometrically, this means that* $O_s(p;\Lambda)$ *acts as* $O(p)$ *on sub-ellipsoids corresponding to axes of the same length of the main ellipsoid, while it acts as a distinct group on the other sub-ellipsoids.*

Remark 4.3.26. *Basically the two groups* $O_s(p;\Lambda)$ *and* $O(p)$ *have either trivial intersection or they coincide.*

Proof. Suppose that $\lambda_1^2 = \dots = \lambda_p^2$, then $\lambda_1^2 = \dots = \lambda_p^2 = 1$ because

$$\lambda_1^2 \times \dots \times \lambda_p^2 = 1$$

and so $\Lambda = Id_{p\times p}$ and so $K^T K = K^T Id_{p\times p} K = K^T \Lambda K = \Lambda = Id_{p\times p}$.

Suppose that there exist two indexes $1 \leq i \neq j \leq n$ such that $\lambda_i^2 \neq \lambda_j^2$. Without loss of generality, we can think that $i = 1$ and $j = 2$. Consider a matrix K of the form

$$K := \begin{pmatrix} 0 & \mu_1 & 0 & \dots & 0 \\ \mu_2 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix}$$

and whose transpose is

$$K^T = \begin{pmatrix} 0 & \mu_2 & 0 & \dots & 0 \\ \mu_1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix}.$$

Now we compute $K^T \Lambda K$ and we get

$$K^T \Lambda K = \begin{pmatrix} \mu_1^2 \lambda_2^2 & 0 & 0 & \dots & 0 \\ 0 & \mu_2^2 \lambda_1^2 & \dots & \dots & 0 \\ 0 & 0 & \lambda_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \lambda_p^2 \end{pmatrix}.$$

If we impose $K^T \Lambda K = \Lambda$, we obtain the conditions:

$$\mu_1^2 \lambda_2^2 = \lambda_1^2 \quad \mu_2^2 \lambda_1^2 = \lambda_2^2.$$

Therefore, for $\mu_1^2 = \frac{\lambda_1^2}{\lambda_2^2}$ and for $\mu_2^2 = \frac{\lambda_2^2}{\lambda_1^2}$, the matrix $K \in O_s(p; \Lambda)$. Since by hypothesis $\lambda_1^2 \neq \lambda_2^2$, then either μ_1 or μ_2 are bigger than one, and so K has at least one entry bigger than 1. This implies that $K \notin O(p)$ and so that $O_s(p; \Lambda) \neq O(p)$, because the two sets differ for at least one element.

Now, $\{Id_{k \times k}\} \in O_s(p; \Lambda)|_{span\{e_{i_1}, \dots e_{i_k}\}} \cap O(p)$, where $e_{i_1}, \dots e_{i_k}$ are eigenvectors of Λ corresponding to different eigenvalues. Now, suppose that there are two eigenvectors e_i, e_j of Λ such that the respective eigenvalues $\lambda_i^2 \neq \lambda_j^2$. Without loss of generality, we can think that $i = 1$ and $j = 2$. Suppose that there exists $K \in O_s(p; \Lambda) \cap O(p)$ with $K \neq Id_{p \times p}$. Then, imposing $K^T \Lambda K = \Lambda$, we have the following system to be solved:

$$\begin{cases} K_{11}^2 \lambda_1^2 + K_{21}^2 \lambda_2^2 = \lambda_1^2 \\ K_{11} K_{12} \lambda_1^2 + K_{21} K_{22} \lambda_2^2 = 0 \\ K_{12}^2 \lambda_1^2 + K_{22}^2 \lambda_2^2 = \lambda_2^2. \end{cases}$$

There are two possible cases $K_{12} = 0$ which implies $K_{21} = 0$ and $K_{22}^2 = K_{11}^2 = 1$. If $K_{12} \neq 0$, then

$$K|_{span\{e_1, e_2\}} = \begin{pmatrix} K_{22} & -\frac{\lambda_2}{\lambda_1} \sqrt{1 - K_{22}^2} \\ \frac{\lambda_1}{\lambda_2} \sqrt{1 - K_{22}^2} & K_{22} \end{pmatrix}.$$

If $K_{22}^2 \neq 1$, then $K|_{span\{e_1, e_2\}}$ must be antisymmetric and so $\lambda_1^2 = \lambda_2^2$, which is absurd. Therefore

$$K|_{span\{e_1, e_2\}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Therefore, all the other entries of the first two columns and rows are zero, otherwise the matrix K cannot be orthogonal. We can repeat the same argument for every pair of eigenvectors of Λ with different eigenvalues. The remaining part of the proof, follows from the first part of the theorem. □

It is easy to find *Maximal Invariant* functions under the group $O_s(p;\Lambda)$.

Proposition 4.3.27. *The function* $g_\Lambda : \mathbb{R}^p \to \mathbb{R}$ *given by* $g_\Lambda(x) = x^T \Lambda x$ *is* Maximal Invariant *under* $(G, \cdot)$ *with* $G = O_s(p;\Lambda)$.

Proof. The function g is invariant under $(O_s(p;\Lambda), \cdot)$, in fact

$$g_\Lambda(Kx) = x^T K^T \Lambda K x = g_\Lambda(x)$$

for all $x \in \mathbb{R}^p$ and for all $K \in O_s(p;\Lambda)$. On the other hand, if $g_\Lambda(x_1) = g_\Lambda(x_2)$ for some $x_1, x_2 \in \mathbb{R}^p$, then there exists $K \in O_s(p;\Lambda)$ such that $x_2 = Kx_1$. This implies that $x_1 \sim x_2$ and so that g_Λ is maximal invariant. □

Remark 4.3.28. *Note that the equivalence relation* $\sim$ *is done through the group*

$$(O_s(p;\Lambda), \cdot)$$

and not through $(O(p), \cdot)$.

The following corollary states that $O_s(p;\Lambda)$ is the invariant group for the $(p-1)$-dimensional ellipsoid $\mathbb{E}_\Lambda^{p-1} := \{x \in \mathbb{R}^p : x^T \Lambda x = 1\}$.

Corollary 4.3.29. *Suppose that* $x \in \mathbb{E}_\Lambda^{p-1}$. *Then, for any* $K \in O_s(p;\Lambda)$, *we have that* $Kx \in \mathbb{E}_\Lambda^{p-1}$, *namely* $\mathbb{E}_\Lambda^{p-1}$ *is invariant under the action of* $O_s(p;\Lambda)$.

Proof. It is a simple application of the previous proposition. □

4.3.6.2 Explicit form of Stretched Orthogonal Matrices in dimension $p = 2$

In this subsection, we provide the explicit form of matrices belonging to $O_s(2, \Lambda_2)$.

Theorem 4.3.30. *Suppose that* $K \in O_s(2, \Lambda_2)$. *Then, there exists* $\theta \in [0, 2\pi)$ *such that*

$$K = K_+ := \begin{pmatrix} \cos(\theta) & -\left|\frac{\lambda_2}{\lambda_1}\right| \sin(\theta) \\ \left|\frac{\lambda_1}{\lambda_2}\right| \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

or

$$K = K_- := \begin{pmatrix} \cos(\theta) & \left|\frac{\lambda_2}{\lambda_1}\right| \sin(\theta) \\ \left|\frac{\lambda_1}{\lambda_2}\right| \sin(\theta) & -\cos(\theta) \end{pmatrix}.$$

Remark 4.3.31. *It is a straightforward computation to verify that $K_+J = K_-$, where the matrix J represents the reflection with respect to the y-axis and takes the form:*

$$J := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Proof. Consider a matrix K of the form

$$K := \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$$

and impose the condition:

$$K^T \Lambda K = \Lambda$$

with

$$\Lambda := \begin{pmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{pmatrix}.$$

The condition

$$K^T \Lambda K = \Lambda$$

reduces to the following system of couple equations:

$$\begin{cases} K_{11}^2\lambda_1^2 + K_{21}^2\lambda_2^2 = \lambda_1^2 \\ K_{11}K_{12}\lambda_1^2 + K_{21}K_{22}\lambda_2^2 = 0 \\ K_{12}^2\lambda_1^2 + K_{22}^2\lambda_2^2 = \lambda_2^2. \end{cases}$$

Suppose that $K_{12} = 0$, then $K_{21} = 0$ and $K_{22}^2 = K_{11}^2 = 1$ and so the only possible solutions are

$$K := \begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}.$$

If $K_{12} \neq 0$, then one has

$$\begin{cases} K_{11} = -\frac{K_{21}K_{22}\lambda_2^2}{K_{12}\lambda_1^2} \\ K_{11}^2\lambda_1^2 + K_{21}^2\lambda_2^2 = \lambda_1^2 \\ K_{12}^2\lambda_1^2 + K_{22}^2\lambda_2^2 = \lambda_2^2. \end{cases}$$

Now, sequentially substituting and using the fact that the function $f(x) = \frac{x}{\sqrt{1-x^2}}$ is monotone in its domain of definition to obtain that $K_{22} = \pm K_{11}$, we get that $K_{22} = \pm K_{11}$, $K_{12} = \pm \left|\frac{\lambda_2}{\lambda_1}\right| \sin(\theta)$ and $K_{21} = \pm \left|\frac{\lambda_1}{\lambda_2}\right| \sin(\theta)$ with $K_{21}K_{12} > 0$ if $K_{11}K_{22} < 0$ and $K_{21}K_{12} < 0$ if $K_{11}K_{22} > 0$. Since $-1 \leq K_{11} \leq 1$ and by choosing $K_{11} = \cos(\theta)$ for some

$\theta \in [0, 2\pi]$, we get that the only possible solutions K are of the form:

$$K = K_+ := \begin{pmatrix} \cos(\theta) & -\left|\frac{\lambda_2}{\lambda_1}\right| \sin(\theta) \\ \left|\frac{\lambda_1}{\lambda_2}\right| \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

or

$$K = K_- := \begin{pmatrix} \cos(\theta) & \left|\frac{\lambda_2}{\lambda_1}\right| \sin(\theta) \\ \left|\frac{\lambda_1}{\lambda_2}\right| \sin(\theta) & -\cos(\theta) \end{pmatrix}.$$

□

Remark 4.3.32. *Note that all the topological and differential properties of $O_s(p;\Lambda)$ are the same as the ones of $O(p)$, since the two groups are diffeomorphic. In fact, the simple rescaling of the axes, $(x_1,\dots,x_p) \mapsto (x_1/\lambda_1,\dots,x_p/\lambda_p)$ the ellipsoid is mapped to the sphere and any $O_s(p;\Lambda)$ is mapped to $O(p)$ by $A \mapsto \Lambda^{-1/2}A$.*

Remark 4.3.33. *A similar analysis can lead to the explicit formula for* Stretched Orthogonal Matrices *in higher dimensions.*

4.3.7 Proof of the main theorems

In this subsection, we prove the characterization of elliptical distributions and of their marginals and conditionals by means of the invariance under the group $O_s(n;\Lambda)$.

4.3.7.1 Characterizations of Elliptical Distributions: proof of Theorem 4.3.3

In this subsubsection, we prove the characterization of *Elliptical Distributions* by means of *Stretched Orthogonal Matrices*, given in Theorem 4.3.3.

We start by proving that the first two conditions of Theorem 4.3.3 are equivalent. Suppose that $\mathbf{X} \stackrel{d}{=} K\mathbf{X}$ for every $K \in O_s(p;\Lambda)$. Note that:

$$\phi_{\Lambda,K\mathbf{X}}(\mathbf{t}) = E\left[e^{i\mathbf{t}^T K\mathbf{x}}\right] = E\left[e^{i(K^T\mathbf{t})^T\mathbf{X}}\right] = \phi_{\Lambda,\mathbf{X}}(K^T\mathbf{t}).$$

But since $\mathbf{X} \stackrel{d}{=} K\mathbf{X}$, then $\phi_{\Lambda,K\mathbf{X}}(\mathbf{t}) = \phi_{\Lambda,\mathbf{X}}(\mathbf{t})$ and so we have that $\phi_{\Lambda,\mathbf{X}}(K^T\mathbf{t}) = \phi_{\Lambda,\mathbf{X}}(\mathbf{t})$.

Now, suppose that $\phi_{\Lambda,\mathbf{X}}(K^T\mathbf{t}) = \phi_{\Lambda,\mathbf{X}}(\mathbf{t})$. By the same chain of equalities, we get $\phi_{K\mathbf{X}}(\mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t})$. But, by the Characterization Theorem of Distributions through their Characteristic Function, we get: $\mathbf{X} \stackrel{d}{=} K\mathbf{X}$. This implies that the first two conditions are equivalent.

To enlighten the notation, from now on, we will not specify the random variable $\mathbf{X}$ for the Characteristic Function and the Maximal Invariant Function, when we do not think there is risk of confusion.

Now, we prove that the second and third conditions are equivalent. Suppose that there exists a function $\psi_\Lambda(\cdot)$ such that $\psi_\Lambda(\mathbf{t}^T\Lambda\mathbf{t}) = \phi(\mathbf{t})$. Then

$$\phi_\Lambda(K^T\mathbf{t}) = \psi((K^T\mathbf{t})^T\Lambda(K^T\mathbf{t})) = \psi(\mathbf{t}^T K\Lambda K^T\mathbf{t}) = \psi(\mathbf{t}^T\Lambda^T\mathbf{t}) = \psi(\mathbf{t}^T\Lambda\mathbf{t}) = \phi_\Lambda(\mathbf{t})$$

and so $\phi_\Lambda(K^T\mathbf{t}) = \phi_\Lambda(\mathbf{t})$.

Now, suppose that $\phi_\Lambda(K^T\mathbf{t}) = \phi_\Lambda(\mathbf{t})$. Then, the function $f_\Lambda(\mathbf{t}) := \mathbf{t}^T\Lambda\mathbf{t}$ is maximal invariant under the action of $(O_s(p;\Lambda),\cdot)$. Therefore, ϕ_Λ must be a function of $f_\Lambda(\mathbf{t}) = t^T\Lambda t$. This implies that there exists ν such that $\phi_\Lambda = \nu \circ f_\Lambda$. The function $\phi_\Lambda = \nu \circ f_\Lambda$ concludes the proof.

Now, we prove that the first and forth conditions are equivalent. Suppose that

$$f_{\mathbf{X}}(\mathbf{x}) = c_p det(\Lambda)^{-1/2} e\left(\mathbf{x}^T\Lambda\mathbf{x}\right),$$

for some e and c_p. Then

$$f_{\mathbf{X}}(K\mathbf{x}) = c_p det(\Lambda)^{-1/2} e\left((K\mathbf{x})^T\Lambda K\mathbf{x}\right) =$$

$$c_p det(\Lambda)^{-1/2} e\left(\mathbf{x}^T K^T\Lambda K\mathbf{x}\right) = c_p det(\Lambda)^{-1/2} e\left(\mathbf{x}^T\Lambda\mathbf{x}\right) = f_{\mathbf{X}}(\mathbf{x}).$$

By the change of variables formula we have:

$$f_{K\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(K^T\mathbf{x})|det(K)| = f_{\mathbf{X}}(K^T\mathbf{x})$$

and so that $f_{\mathbf{X}}(\mathbf{x}) = f_{K\mathbf{X}}(\mathbf{x})$ and that $\mathbf{X} \overset{d}{=} K\mathbf{X}$ for every $K \in O_s(p;\Lambda)$. On the other side, if $\mathbf{X} \overset{d}{=} K\mathbf{X}$, then with the same argument used backwards $f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(K\mathbf{x})$ and so the density $f_{\mathbf{X}}(\mathbf{x})$ is invariant under the action of $O_s(p;\Lambda)$. But $O_s(p;\Lambda)$ has the function $f_\Lambda(t) = \mathbf{t}^T\Lambda\mathbf{t}$ as a maximal invariant function and so $f_{\mathbf{X}}$ must be a function of ψ

(Proposition 4.3.16) and so it must take the form:

$$f_{\mathbf{X}}(\mathbf{x}) = c_p det(\Lambda)^{-1/2} e\left(\mathbf{x}^T \Lambda \mathbf{x}\right),$$

with c_p a normalizing constant dependent just on the dimension p and the function e and the function e is independent of p.

Example 4.3.34. *Well known examples of Elliptical distributions are the* Multivariate Normal Distribution*,* Multivariate t-Distribution*,* Mixtures of Multivariate Normal Distributions*. Slightly less known is the explicit formula for The* Uniform Distribution on the Ellipsoid $X \sim Unif(\mathbf{E}(\Lambda))$ *is given by*

$$f_X(x) = \frac{1}{S}, \qquad x \in \mathbf{E}(\Lambda),$$

where S is given by:

$$S = 2\pi c^2 + \frac{2\pi ab}{\sin\phi}\left(E(\phi,k)\,\sin^2\phi + F(\phi,k)\,\cos^2\phi\right).$$

Here

$$\cos\phi = \frac{c}{a}, \qquad k^2 = \frac{a^2(b^2-c^2)}{b^2(a^2-c^2)}, \qquad a \geq b \geq c$$

and

$$F(\phi,k) := \int_0^\phi \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}}$$

and

$$E(\phi,k) := \int_0^\phi \sqrt{1-k^2\sin^2\theta}\,d\theta$$

are incomplete elliptic integrals of the first and second kind, respectively. The case of the sphere ($a = b = c$ which gives $\phi = 0$) is well defined in the limit $\phi \to 0$. Note that the surface area of an ellipsoid of revolution may be expressed in terms of elementary functions:

$$S_1 = 2\pi a^2\left(1 + \frac{1-e^2}{e}\tanh^{-1} e\right) \quad \text{where} \quad e^2 = 1 - \frac{c^2}{a^2} \quad (c < a),$$

$$S_2 = 2\pi a^2\left(1 + \frac{c}{ae}\sin^{-1} e\right) \qquad \text{where} \quad e^2 = 1 - \frac{a^2}{c^2} \quad (c > a).$$

Note that $S_1 = S_2$. Any $X \sim Unif(\mathbf{E}(\Lambda))$ is invariant under the action of $O_s(p;\Lambda)$.

4.3.7.2 Marginals and Conditional Distributions: proof of Theorem 4.3.6

This subsubsection is devoted to the proof of Theorem 4.3.6.

It is well known that all marginals of Λ-elliptical distributions are elliptical with the same functional form (see [90] page 34 or Subsection 4.3.5 above). Note that, by Theorem 4.3.3, $\phi_\Lambda(\mathbf{t}) = \psi\left(\sum_{i=1}^n \lambda_i^2 t_i^2\right)$ for some ψ. If we choose $\mathbf{t} = (\mathbf{t}_1^T, \mathbf{t}_2^T)^T$, with $\mathbf{t}_2 = \mathbf{0}$, then we obtain the characteristic function of the vector $\mathbf{Y}_1$. Again, by Theorem 4.3.6, we have that there exists $\tilde{\psi}$ such that

$$\phi_1(\mathbf{t}_1) := E[e^{i\mathbf{t_1}^T\mathbf{Y}_1}] = \tilde{\psi}\left(\sum_{i=1}^{p_1} \lambda_i^2 t_i^2\right) =$$

$$\psi\left(\frac{\sum_{i=1}^{p_1} \lambda_i^2 t_i^2}{\prod_{i=1}^{p_1} \lambda_i^2}\right)$$

with $\tilde{\psi}(z) =: \psi\left(\frac{z}{\prod_{i=1}^{p_1} \lambda_i^2}\right)$. Therefore, $\mathbf{Y}_1 \overset{\mathrm{d}}{=} K\mathbf{Y}_1$ for every $K \in O_s(p_1; \Lambda_{p_1})$.

Now

$$\tilde{\Lambda}_{p_1} := diag\left(\frac{\lambda_i^2}{\prod_{i=1}^{p_1} \lambda_i^2}\right)_{i=1,\dots,p_1}$$

is a matrix which is allowed in our definition of *Stretched Orthogonal Matrices*, therefore for every $K \in O_s(p_1; \Lambda_{p_1})$, we have that $K^T \tilde{\Lambda}_{p_1} K = \tilde{\Lambda}$ and so $K^T \Lambda_{p_1} K = \Lambda$. This implies that the group $O_s(p_1; \Lambda_{p_1})$ which leaves invariant $\mathbf{Y}_1$ is the subgroup of $O_s(p_1; \Lambda_{p_1})$, which does not move the remaining components $\mathbf{Y}_2$.

For what concerns the conditional distribution part, that is a consequence of Theorem 1.5.4 from Muirhead [90] (see also Subsection 4.3.5) and the proof just developed on the marginal distributions case.

4.3.8 Appendix

In this subsubsection, we collect some further properties of elliptical distributions, that are well known for spherical distributions and that can be proved with simple modifications in the spherical case.

4.3.8.1 Polar Representation

We can represent an elliptically distributed random vector into polar coordinates.

Theorem 4.3.35. *Suppose that* $\mathbf{Y} \sim E(\Lambda)$ *for some* $\Lambda = diag(\lambda_1^2, \ldots, \lambda_p^2) \in Mat_{p\times p}$ *and has density*

$$f_{\mathbf{Y}}(\mathbf{y}) = c_p det(\Lambda)^{-1/2} e\left(\mathbf{y}^T \Lambda \mathbf{y}\right),$$

with c_p *a normalizing constant dependent just on the dimension* p *and the function* e *is independent of* p*. Suppose to perform the change of variables:*

$$\begin{cases} Y_1 = \lambda_1 r \sin(\theta_1)\sin(\theta_2)\ldots\sin(\theta_{p-2})\sin(\theta_{p-1}) \\ Y_2 = \lambda_2 r \sin(\theta_1)\sin(\theta_2)\ldots\sin(\theta_{p-2})\cos(\theta_{p-1}) \\ Y_3 = \lambda_3 r \sin(\theta_1)\sin(\theta_2)\ldots\cos(\theta_{p-2}) \\ \vdots \\ Y_{p-1} = \lambda_{p-1} r \sin(\theta_1)\sin(\theta_2) \\ Y_p = \lambda_p r\cos(\theta_1), \end{cases}$$

with $r > 0, 0 < \theta_i \leq \pi, i = 1, \ldots p-2, 0 < \theta_{p-1} \leq 2\pi$*. Then, the following are true.*

- *The random variables* $r, \theta_1, \ldots, \theta_{p-1}$ *are independent.*
- *The distributions of* $\theta_1, \ldots, \theta_{p-1}$ *are the same for all* $\mathbf{Y}$*.*
- *The marginal distributions of the* θ_k*'s are proportional to* $\sin^{p-1-k}(\theta_k)$ *for* $k = 1, \ldots, p-1$*.*
- *The random variable* $r^2 := Y^T \Lambda^{-1} Y$ *has density function given by:*

$$f_{r^2}(w) = \frac{c_p \pi^{p/2}}{\Gamma(\frac{p}{2})} w^{p/2-1} e(w), \qquad w > 0.$$

Proof. The proof dips most of the steps from Theorem 1.5.5 page 55 from [90]. By the change of variables $\mathbf{Y} = \Lambda^{1/2}\mathbf{X}$, then $\mathbf{X}$ is spherically distributed with density:

$$f_{\mathbf{X}}(\mathbf{x}) = c_p e\left(\mathbf{x}^T \mathbf{x}\right).$$

It is well know (see for example Theorem 2.1.3 in [90]) that the transformation from $X_1, \ldots, X_p$ to $r, \theta_1, \ldots, \theta_{p-1}$ is

$$J(r, \theta_1, \ldots, \theta_{p-1}) = r^{p-1} \sin^{p-2}(\theta_1) \sin^{p-3}(\theta_2) \ldots \sin(\theta_{p-2}).$$

It follows that the joint density of $r, \theta_1, \ldots, \theta_{p-1}$ is

$$f_{(r,\theta_1,\ldots,\theta_{p-1})} = \frac{c_p}{2} r^{p-2} \sin^{p-2}(\theta_1) \sin^{p-3}(\theta_2) \ldots \sin(\theta_{p-2}) e(r^2).$$

A first simple consequence of this formula is that $r, \theta_1, \dots, \theta_{p-1}$ are independent random variables by *Factorization Theorem* and that the distributions of $\theta_1, \dots, \theta_{p-1}$ are the same for all $\mathbf{Y}$. Compute the marginal distribution of r^2 f_{r^2} by integrating against the angles and get (for $w = r^2$):

$$f_{r^2}(w) = \int_{\theta_1,\dots,\theta_{p-1}} f_{(r,\theta_1,\dots,\theta_{p-1})} = \int_{\theta_1,\dots,\theta_{p-1}} \frac{c_p}{2} r^{p-2} \sin^{p-2}(\theta_1)\sin^{p-3}(\theta_2)\dots\sin(\theta_{p-2})e(r^2)$$

$$= \frac{c_p}{2} r^{p-2} e(r^2) \int_{\theta_1,\dots,\theta_{p-1}} \sin^{p-2}(\theta_1)\sin^{p-3}(\theta_2)\dots\sin(\theta_{p-2}) = \frac{c_p \pi^{p/2}}{\Gamma(\frac{p}{2})} w^{p/2-1} e(w), \quad w > 0.$$

□

A simple consequence of the previous theorem is the following corollary.

Corollary 4.3.36. *Suppose that* $\mathbf{Y} \sim E(\Lambda)$ *for some* $\Lambda = diag(\lambda_1^2, \dots, \lambda_p^2) \in Mat_{p\times p}$ *and has density*

$$f_{\mathbf{Y}}(y) = c_p det(\Lambda)^{-1/2} e\left(\mathbf{y}^T \Lambda \mathbf{y}\right),$$

with c_p *a normalizing constant dependent just on the dimension* p *and the function* e *and the function* e *is independent of* p*. Suppose to have the change of variables:*

$$\begin{cases} Y_1 = \lambda_1 r \sin(\theta_1)\sin(\theta_2)\dots\sin(\theta_{p-2})\sin(\theta_{p-1}) \\ Y_2 = \lambda_2 r \sin(\theta_1)\sin(\theta_2)\dots\sin(\theta_{p-2})\cos(\theta_{p-1}) \\ Y_3 = \lambda_3 r \sin(\theta_1)\sin(\theta_2)\dots\cos(\theta_{p-2}) \\ \vdots \\ Y_{p-1} = \lambda_{p-1} r \sin(\theta_1)\sin(\theta_2) \\ Y_p = \lambda_p r\cos(\theta_1), \end{cases}$$

with $r > 0, 0 < \theta_i \leq \pi, i = 1, \dots p-2, 0 < \theta_{p-1} \leq 2\pi$*. Then, the* $\sin^2\theta_k$*'s have the* Beta Distribution*, for* $k = 1, \dots, p-2$*:*

$$f_{\sin^2\theta_k}(y) = \frac{y^{\frac{p-k}{2}-1}(1-y)^{\frac{1}{2}-1}}{B(\frac{1}{2}, \frac{p-k}{2})}, \qquad 0 \leq \theta_k < \pi, k = 1, \dots, p-2$$

while $\theta_{p-1} \sim Unif(0, 2\pi)$*.*

4.3.8.2 Angle-Radius Theorems

In this subsubsubsection, we discuss some independence properties of the angle and radius random variables, which are consequences of the polar representation.

First of all, we have that the radius and angles are independent.

Theorem 4.3.37. *Suppose that* $\mathbf{Y}$ *is a* Λ*-Elliptical Distribution such that* $P(\mathbf{Y}=\mathbf{0})=0$*. Define* $\Omega_\Lambda(\mathbf{Y}) := \frac{\mathbf{Y}}{\mathbf{Y}^T\Lambda^{-1}\mathbf{Y}}$ *and* $r_\Lambda := \mathbf{Y}^T\Lambda^{-1}\mathbf{Y}$*. Then:*

- $\Omega_\Lambda(\mathbf{Y})$ *is uniformly distributed on* $E(\Lambda)$*.*
- $\Omega_\Lambda(\mathbf{Y})$ *and* $r_\Lambda(\mathbf{Y})$ *are independent.*

Remark 4.3.38. *The condition* $P(\mathbf{Y}=\mathbf{0})=0$ *is a technical condition, assumed to ensure that all the quantities involved are well defined with probability one.*

Proof. Consider $K \in O_s(p;\Lambda)$, then:

$$\Omega_\Lambda(K\mathbf{Y}) = \frac{K\mathbf{Y}}{(K\mathbf{Y})^T\Lambda^{-1}K\mathbf{Y}} = K\frac{\mathbf{Y}}{\mathbf{Y}^TK^T\Lambda^{-1}K\mathbf{Y}} = K\frac{\mathbf{Y}}{\mathbf{Y}^T\Lambda^{-1}\mathbf{Y}} = K\Omega_\Lambda(\mathbf{Y}),$$

since whenever $K \in O_s(p;\Lambda)$, also $K^{-1} \in O_s(p;\Lambda)$. Therefore $\Omega_\Lambda(K\mathbf{Y}) \sim K\Omega_\Lambda(\mathbf{Y})$. Since $\mathbf{Y} \sim E(\Lambda)$, then $\mathbf{Y} \sim K\mathbf{Y}$ and so $\Omega_\Lambda(K\mathbf{Y}) \sim \Omega_\Lambda(\mathbf{Y})$. Therefore $K\Omega_\Lambda(\mathbf{Y}) \sim \Omega_\Lambda(\mathbf{Y})$ and hence by Theorem 4.3.47, $\Omega_\Lambda(\mathbf{Y})$ must be the uniform distribution on $E(\Lambda)$.
Define:

$$\mu(B) := P(\Omega_\Lambda(\mathbf{Y}) \in B | r \in C),$$

where B is a Borel set in $E(\Lambda)$ and C is a Borel set such that $P(r \in C) \neq 0$. The measure μ is invariant under the action of $O_s(p;\Lambda)$, since:

$$\mu(KB) := P(\Omega_\Lambda(\mathbf{Y}) \in KB | r \in C) = P(K^{-1}\Omega_\Lambda(\mathbf{Y}) \in B | r \in C)$$

$$= P(\Omega_\Lambda(K^{-1}\mathbf{Y}) \in B | r \in C) = P(\Omega_\Lambda(\mathbf{Y}) \in B | r \in C) = \mu(B).$$

Therefore $\mu(KB) = \mu(B)$ for any $K \in O_s(p;\Lambda)$ and so the measure μ is invariant under the action of $O_s(p;\Lambda)$ and so it must be the uniform distribution on $E(\Lambda)$. Therefore,

$$P(\Omega_\Lambda(\mathbf{Y}) \in B | r \in C) = \mu(B) = P(\Omega_\Lambda(\mathbf{Y}) \in B),$$

which means that $P(\Omega_\Lambda(\mathbf{Y}) \in B | r \in C) = P(\Omega_\Lambda(\mathbf{Y}) \in B)$ and so r_Λ and Ω_Λ are independent. □

We also have a characterization of t-distribution and *Beta*-distribution, similar to the one using spherical distributions (see [90] Theorem 1.5.6 and Theorem 1.5.7 page 38).

Theorem 4.3.39. *Suppose that* $\mathbf{Y} \sim E(\Lambda)$ *is a* $p \times 1$ *random vector, such that* $P(\mathbf{Y} = \mathbf{0}) = 0$. *Then, the following are true:*

- $W := \frac{\alpha^T \Lambda^{-1/2}\mathbf{Y}}{\|\Lambda^{-1/2}\mathbf{Y}\|}$, *for some* $\alpha^T\alpha = 1$, $\alpha \in \mathbb{R}^p$, *then*

$$T = (p-1)^{\frac{1}{2}} \frac{W}{(1-W^2)^{\frac{1}{2}}}$$

has the t_{p-1} *distribution.*

- *If* $A \in Sym^+_{p\times p}$ *idempotent of rank* k, *then*

$$Z = \frac{\mathbf{Y}^T \Lambda^{-1/2} A \Lambda^{-1/2}\mathbf{Y}}{\|\Lambda^{-1/2}\mathbf{Y}\|^2}$$

has the Beta Distribution $Z \sim Beta\left(\frac{k}{2}, \frac{p-k}{2}\right)$.

Proof. Consider the random variable:

$$T = (p-1)^{\frac{1}{2}} \frac{W}{(1-W^2)^{\frac{1}{2}}}.$$

Note that T is a function of Ω_Λ which is uniformly distribution on $E(\Lambda)$ as soon as T is Λ-Elliptical by Theorem 4.3.37 and for every α as in the hypotheses. Therefore, we have freedom in the choice of the distribution of Y. In the meantime:

$$T = (p-1)^{\frac{1}{2}} \frac{W}{(1-W^2)^{\frac{1}{2}}} = (p-1)^{\frac{1}{2}} \frac{\frac{\alpha^T\Lambda^{-1/2}\mathbf{Y}}{\|\Lambda^{-1/2}\mathbf{Y}\|}}{(1-\left(\frac{\alpha^T\Lambda^{-1/2}\mathbf{Y}}{\|\Lambda^{-1/2}\mathbf{Y}\|}\right)^2)^{\frac{1}{2}}}$$

$$= (p-1)^{\frac{1}{2}} \frac{\alpha^T\Lambda^{-1/2}\mathbf{Y}}{(\|\Lambda^{-1/2}\mathbf{Y}\|^2 - \left(\alpha^T\Lambda^{-1/2}\mathbf{Y}\right)^2)^{\frac{1}{2}}}$$

$$= (p-1)^{\frac{1}{2}} \frac{\sum_{k=1}^p \frac{\alpha_k Y_k}{\lambda_k^2}}{\left[\sum_{j=1}^p \left(\frac{Y_j}{\lambda_j^2}\right)^2 - \left(\sum_{i=1}^p \frac{\alpha_i Y_i}{\lambda_i^2}\right)^2\right]^{\frac{1}{2}}}.$$

Since we are allowed to choose our favourite α by Theorem 4.3.37, we choose α such that

$\alpha_k = 0$ for every $k = 2, \dots, p$ and so we obtain:

$$Y = (p-1)^{\frac{1}{2}} \frac{\frac{\alpha_1 Y_1}{\lambda_1^2}}{\left[\sum_{j=1}^p \left(\frac{Y_j}{\lambda_j^2}\right)^2 - \left(\frac{\alpha_1 Y_1}{\lambda_1^2}\right)^2\right]^{\frac{1}{2}}} = (p-1)^{\frac{1}{2}} \frac{\frac{\alpha_1 Y_1}{\lambda_1^2}}{\left[\sum_{j=2}^p \left(\frac{Y_j}{\lambda_j^2}\right)^2 + \left(\frac{Y_1}{\lambda_1^2}\right)^2 - \left(\frac{\alpha_1 Y_1}{\lambda_1^2}\right)^2\right]^{\frac{1}{2}}}.$$

Therefore, if we choose $\alpha_1 = 1$ and $Y_k \sim N(0, \lambda_k^2)$ for $k = 1, \dots, p$ we get that

$$Y = (p-1)^{\frac{1}{2}} \frac{Z_1}{\sqrt{Y_{p-1}^2}},$$

with

$$Z_1 := \frac{Y_1}{\lambda_1^2} \sim N(0,1); \quad Y_{p-1}^2 := \sum_{j=2}^p \left(\frac{Y_j}{\lambda_j^2}\right)^2 \sim \chi^2(p-1).$$

Therefore, since Z_1 and Y_{p-1}^2 are independent and by definition of t-distribution $Y \sim t_{p-1}$.

Similarly,

$$Z = \frac{\mathbf{Y}^T \Lambda^{-1/2} A \Lambda^{-1/2} \mathbf{Y}}{\|\Lambda^{-1/2}\mathbf{Y}\|^2}$$

with $A \in Sym_{p\times p}^+$ idempotent of rank k. Again, since

$$Z = \Omega_\Lambda(\mathbf{Y}) A \Omega_\Lambda(\mathbf{Y}),$$

we can assume without loss of generality that $Y \sim \mathcal{MVN}(\mathbf{0}, \Lambda)$. Therefore, there exists $K \in O_s(p; \Lambda)$ such that

$$K^T \Lambda^{-1/2} A \Lambda^{-1/2} K = \begin{bmatrix} Id_k & 0 \\ 0 & 0 \end{bmatrix}.$$

Define $\mathbf{U} := K\Lambda^{-1/2}\mathbf{Y}$. We have:

$$Z = \frac{\sum_{j=1}^k U_j^2}{\sum_{j=1}^n U_j^2},$$

where $\sum_{j=1}^k U_j^2 \sim \chi_k^2$ and $\sum_{j=K+1}^n U_j^2 \sim \chi_{n-k}^2$ are independent. This implies that

$$Z \sim Beta\left(\frac{k}{2}, \frac{p-k}{2}\right).$$

□

4.3.8.3 The Haar Measure on $O_s(p;\Lambda)$

In this subsubsection, we construct explicitly the *Haar measure* for $O_s(p;\Lambda)$. We recall the precise definition of *Haar measure*.

Definition 4.3.40. *Consider a locally compact Hausdorff topological group $(G,\cdot)$, endowed with the Borel σ-algebra $\mathcal{B}$. We define the* left translate *of $S \in \mathcal{B}$ as follows:*

$$gS = \{g.s \,:\, s \in S\}.$$

A measure μ on $\mathcal{B}$ is called left translation invariant *if for every $S \in \mathcal{B}$ and every $g \in G$ we have:*

$$\mu(gS) = \mu(S).$$

A well known definition-theorem due to Haar gives the definition of Haar measure.

Theorem 4.3.41. *The* Haar measure *is the unique, up to a positive multiplicative constant, countably additive, non-trivial measure μ on $\mathcal{B}$ satisfying the following properties:*

- *The measure μ is left-translation-invariant.*
- *The measure μ is* finite *on every compact set, namely $\mu(K) < +\infty$ for every K compact with respect to the topology of the group.*
- *The measure is* outer regular *on $\mathcal{B}$:*

$$\mu(E) = \inf\{\mu(U) : E \subseteq U, U \text{ open}\}.$$

- *The measure μ is* inner regular *on open sets:*

$$\mu(E) = \sup\{\mu(K) : K \subseteq E, K \text{ compact}\}.$$

Now, we can pass to our specific case. Consider any $K \in O_s(p;\Lambda)$. We have

$$K^T \Lambda K = \Lambda.$$

Take the exterior derivative of this equality. Since Λ is constant, we get:

$$dK^T \Lambda K + K^T \Lambda dK = 0_{p\times p},$$

which implies that

$$dK^T \Lambda K = -(dK^T \Lambda K)^T$$

and so that $dK^T \Lambda K$ is antisymmetric. Note also that, if we define

$$H := \Lambda^{1/2} K \Lambda^{-1/2},$$

then

$$H^T H = (\Lambda^{1/2} K \Lambda^{-1/2})^T \Lambda^{1/2} K \Lambda^{-1/2} = Id_{p \times p}$$

and so the matrix H is orthogonal. Therefore, with an analogous computation, we get:

$$H^T dH = \Lambda^{-1/2} K^T \Lambda dK \Lambda^{-1/2} := \omega.$$

The differential form ω is invariant under the action of $O_s(p; \Lambda)$. In fact, it is left invariant, since for $P \in O_s(p; \Lambda)$, $K \mapsto PK$ implies $\Lambda^{-1/2} K^T P^T \Lambda P dK \Lambda^{-1/2} = \Lambda^{-1/2} K^T \Lambda dK \Lambda^{-1/2}$. It is also right invariant ($K \mapsto KP^T$) by a similar computation and Theorem 2.16 of [90]. Therefore

$$(\omega) = (\Lambda^{-1/2} K^T \Lambda dK \Lambda^{-1/2}) := \bigwedge_{i<j}^{m} (\Lambda^{-1/2} K^T \Lambda^{1/2})_i (\Lambda^{1/2} dK \Lambda^{-1/2})_j$$

is an invariant differential form under the action of $O_s(p, \Lambda)$.

Therefore, we can define the *Haar Measure* of $O_s(p; \Lambda)$ by

$$\mu(S) := \int_S (\omega)$$

for every $S \subset O_s(p; \Lambda)$. It is easy to see that

$$\mu(S) = \mu(PS) = \mu(SP^T),$$

for any $P \in O_s(p; \Lambda)$. Therefore μ is an invariant measure under the action of $O_s(p; \Lambda)$.

Recall the following theorem (Theorem 2.1.15 and Corollary 2.1.16) from [90].

Theorem 4.3.42.

$$Vol(O(p)) = \frac{2^p \pi^{p^2/2}}{\Gamma_p(\frac{p}{2})}.$$

Using this and that $Vol(O(p)) = Vol(O(p)) det(\Lambda) = Vol(O_s(p; \Lambda))$, we can easily construct a normalized differential form

$$(\omega^*) := \frac{1}{Vol(O_s(p;\Lambda))}(\omega)$$

and so

$$\mu^*(S) := \int_S (\omega^*)$$

an invariant probability measure on $O_s(p;\Lambda)$.

We can give the explicit form of the Haar measure in the case $p = 2$

Example 4.3.43. *Consider the case $p = 2$. In this case, the matrix K takes the form*

$$K = K_+ := \begin{pmatrix} \cos(\theta) & -\left|\frac{\lambda_2}{\lambda_1}\right|\sin(\theta) \\ \left|\frac{\lambda_1}{\lambda_2}\right|\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Therefore

$$K^T \Lambda dK$$

$$= \begin{pmatrix} \cos(\theta) & -\left|\frac{\lambda_2}{\lambda_1}\right|\sin(\theta) \\ \left|\frac{\lambda_1}{\lambda_2}\right|\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{pmatrix} \begin{pmatrix} -\sin(\theta)d\theta & -\left|\frac{\lambda_2}{\lambda_1}\right|\cos(\theta)d\theta \\ \left|\frac{\lambda_1}{\lambda_2}\right|\cos(\theta)d\theta & -\sin(\theta)d\theta \end{pmatrix}.$$

Therefore

$$K^T \Lambda dK = d\theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

This implies that

$$\mu^*_{O_s(2;\Lambda_2)} := \int_S (\omega^*) = \frac{1}{Vol(O_s(2;\Lambda_2))}\int_S (\omega) = \int_{\tilde{S}} d\theta.$$

Note that all this discussion gives us a characterization of $\mu^*_{O_s(p;\Lambda)}$ as the uniform distribution on $O_s(p;\Lambda)$.

Corollary 4.3.44. *The measure $\mu^*_{O_s(p;\Lambda)}$ is the uniform probability distribution on the group $O_s(p;\Lambda)$.*

4.3.8.4 The Spherical and Elliptical Measure

To construct an elliptical measure, we can simply use a diffeomorphism from the sphere to the ellipsoid and pull-back the spherical measure. Here, we use this approach. We start with a definition.

Definition 4.3.45. *The* Spherical Measure $\mu_{\mathbb{S}^n}$ *is the normalized natural Borel measure on* $\mathbb{S}^n$.

This definition does not give a precise way to construct the *Spherical Measure*. There are several ways to do this: by means of the Hausdorff Measure, by Embedding in the Euclidean space or using Lebesgue measure. All these methods define the same measure on $\mathbb{S}^n$. Indeed, by a theorem of Christensen [30], we have the following proposition.

Proposition 4.3.46. *[30] Let M be a locally compact Hausdorff space and u and v two positive uniform measures on M. Then, the exists $c \in \mathbb{R}^+$ such that $u = cv$.*

Since all these measures represent the uniform distribution on $\mathbb{S}^n$, and any two uniformly distributed normalized Borel regular measures on a separable metric space must be the same measure, then all these measures represent the same measure.

The case of the ellipsoid is analogous and therefore, we have the following theorem.

Theorem 4.3.47. *There exists a unique invariant probability distribution $\mu^*_{O_s(p;\Lambda)}$ on $O_s(p;\Lambda)$, which is the uniform distribution on $O_s(p;\Lambda)$ and is given by the formula:*

$$\mu^*_{O_s(p;\Lambda)} := \int_S (\omega^*) = \frac{1}{Vol(O_s(p;\Lambda))} \int_S (\omega)$$

There is a natural relationship between the *Haar Measure* on $O(p)$ and $O_s(p;\Lambda)$ and the *Spherical and Elliptical Measures.*

For any $x \in \mathbb{S}^n$ and any $A \subset \mathbb{S}^n$, we have

$$\mu^*_{O(p)}(\{g \in O(p) \mid g(x) \in A\}) = \sigma^{p-1}(A).$$

Analogously, for any $x \in \mathbb{E}^{p-1}_\Lambda$ and any $A \subset \mathbb{E}^{p-1}_\Lambda$, we have

$$\mu^*_{O_s(p;\Lambda)}(\{h \in O_s(p;\Lambda) \mid h(x) \in A\}) = \sigma^{p-1}_\Lambda(A),$$

where σ^{p-1}_Λ is the *Uniform Measure* on $\mathbb{E}^{p-1}_\Lambda$.

4.3.8.5 The Metric Ground Form

The theory that we discussed in this manuscript can be set in the context of metric ground forms (see [123]).

In his discussion of Classical Groups, H. Weyl [123] considers arbitrary non-degenerate quadratic forms, as

$$Q(\mathbf{x}) = \sum_{i=1}^{m} \sum_{j=1}^{m} q_{ij} x_i x_j,$$

with $q_{ij=q_{ji}}$. In the highest level of generality q_{ij} belong to an arbitrary field $\mathbb{K}$ of zero characteristic for every $i, j = 1, \ldots, m$.

The linear transformations L such that for $Q = [q_{ij}]_{i,j=1}^{m}$, we have

$$L^T Q L = Q$$

form a sort of "orthogonal group" $O_G(m)$.

Our case correspond to $Q = \Lambda$, $G = O_s(m, \Lambda_m)$ and $\mathbb{K} = \mathbb{R}$.

Weyl does not discuss the applications to distribution theory of his characterizations, but one can generalize our constructions using Weyl's theory and characterize more general group invariant probability distributions. However, we find that it is good to have explicitly discussed the case of the broadly used elliptical distributions with specific interest.

4.4 The Direct Product of Circle Groups

The circle group, often denoted by $(\mathbb{T}, \cdot)$, since it can be identified as the 1-torus is the multiplicative subgroup of the complex plane $\mathbb{C}$ of all complex numbers with absolute value 1:

$$\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}.$$

It is an Abelian, since $\mathbb{C}$ is Abelian.

The circle group is isomorphic to other important Lie Groups. We have

$$\mathbb{T} \simeq \mathbb{S}^1 \simeq \mathbb{U}(1) \simeq \mathbb{R} \backslash \mathbb{Z} \simeq \mathbb{SO}(2).$$

Here $\mathbb{U}(1)$ is the set of all unitary matrices, $\mathbb{R} \backslash \mathbb{Z}$ is the quotient space of the real line with respect to integer translations and $\mathbb{SO}(2)$ is the set of special orthogonal matrices of dimension 2.

Motivated by the applications to molecular biology, in the following treatment, we will be concentrated on the more general product space $\mathbb{T}^2$ which corresponds to the 2-torus.

4.5 On Geometric Probability Distributions on the Torus with Applications to Molecular Biology

In this section, we propose a family of probability distributions, alternative to the *von Mises* family, that we call *Inverse Stereographic Normal Distributions*, which are counterparts of the *Gaussian Distribution* on $\mathbb{S}^1$ (univariate) and on $\mathbb{S}^n$ and $\mathbb{T}^n$ (multivariate). We discuss some key properties of the model, such as multimodality and closure with respect to marginalizing and conditioning. We compare this family of distributions to the *von Mises*' family. We then discuss some inferential problems and introduce a notion of moments which is natural for inverse stereographic distributions, revisit a version of the Central Limit Theorem in this context and construct point estimators, confidence intervals and hypothesis tests. Finally, we conclude with some applications to molecular biology and some examples, as to how to use these models to approximate parameters of the *von Mises* distributions in certain particular cases. This study is motivated by the *Protein Folding Problem* and by the fact that a large number of proteins involved in the DNA-metabolism assume a toroidal shape with some amorphous regions.

4.5.1 Introduction and Motivation

In the recent years, statisticians paid increasing attention on random variables taking values on manifolds, since probability distributions on manifolds have key applications and also propose stimulating theoretical challenges. The theory of statistical distributions on manifolds represents the natural link between several theoretical research areas including topology, differential geometry, analysis, probability theory, data science and statistical inference. We refer to [66] and [83] for an extended list of references in this research area.

Amongst the most important statistical problems on manifolds, those arising from the analysis of circular data, spherical data and toroidal data play a fundamental role. Motivated by important applications in molecular biology, we will concentrate mostly on toroidal data, but we will discuss some properties of distributions on the circle and the sphere as well.

More precisely, the motivation behind our study relates to the *Protein Folding Problem*

(PFP), which is one of the major open problems in biochemistry. Protein folding is the physical process that a protein chain undertakes before reaching its final three dimensional shape (conformation). The shape of the protein ensures that the protein does its job properly, while a misfolding can be the cause of several neurodegenerative diseases and other types of diseases as well [106].

Since the physical process underlying protein folding is complicated, there has been limited success in predicting the final folded structure of a protein from its amino acid sequence. A better understanding of this would definitely have clinical impact and result in the design of efficient drug molecules for the cure of the diseases mentioned above.

The PFP can be divided into two parts: the static problem and the dynamic problem. As explained in [41], the static problem is to predict the active conformation of the protein given only its amino acid sequence, while the dynamic problem consists in rationalizing the folding process as a complex physical interaction. Both these problems have a precise mathematical formulation.

Some regions of the conformation of a protein may look amorphous and so require random models and probability distributions to be described properly. A large number of proteins involved in DNA metabolism with different evolutionary origin and catalyzing different reactions assume a toroidal shape [60], [61]. It has been argued that the preference towards the toroidal form may be due to special advantages in the DNA-binding [60], [61]. A long list of proteins share the toroidal form [60], [61].

The importance of the PFP and the DNA-binding process motivated statisticians to find appropriate statistical models that describe these phenomena and then to study the properties of these models which have circular, spherical or toroidal symmetries theoretically as in applications.

The most famous circular distribution is the *von Mises distribution*. We say that a random variable Θ, taking values in $\mathbb{S}^1$ is distributed as a *Von Mises Random Variable* $\Theta \sim VM(\mu, \kappa)$, if its probability density (pdf) is given by

$$f(\theta \mid \mu, \kappa) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)},$$

where $\kappa > 0$, $\mu \in [-\pi, \pi)$ and $\theta \in [-\pi, \pi)$. Here,

$$I_0(z) := \frac{1}{2\pi i} \oint e^{(z/2)(t+1/t)} t^{-1} dt$$

is the *Modified Bessel Function of the First Kind of Order* 0 (the contour encloses the origin and it is run counterclockwise). The parameters $\mu \in [-\pi, \pi)$ and $\kappa > 0$ represent the *location* and *concentration*, respectively. One of the reasons for the popularity of this distribution is that it approaches the normal distribution on the real line in the high-concentration limit ($\kappa \to +\infty$).

The interest on the theory as well as practical applications motivated researchers to look for a good higher dimensional analogue of the one dimensional *von Mises Distribution.* The *Bivariate Von Mises Distribution* (BVM) is a probability distribution describing a two dimensional random vector, taking values on a torus $\mathbb{T}^2 := \mathbb{S}^1 \times \mathbb{S}^1$. It aims at representing an analogue on the torus of the *Bivariate Normal Distribution* (BVN). The *Full Bivariate Von Mises Distribution* (FBVM) was first proposed by Mardia [79]. Some of its variants are currently used in the field of *bioinformatics* to formulate probabilistic models of protein structure. See also [80], [81], [82], [83], [84], [85], [86], [87] and [118].

We say that a random variable $\mathbf{\Theta} = (\Theta_1, \Theta_2)$, taking values in $\mathbb{T}^2$, is distributed as a *Bivariate Von Mises Random Variable* $\mathbf{\Theta} \sim BVM(\mu, \nu, \mathbf{A}, \kappa_1, \kappa_2)$, if its pdf is given by the following function:

$$f(\phi, \psi) =$$

$$\frac{\exp[\kappa_1 \cos(\phi - \mu) + \kappa_2 \cos(\psi - \nu) + (\cos(\phi - \mu), \sin(\phi - \mu))\mathbf{A}(\cos(\psi - \nu), \sin(\psi - \nu))^T]}{\mathcal{Z}},$$

for the angles $\phi, \psi \in [-\pi, +\pi)$. Here, μ and ν are the location parameters for ϕ and ψ, respectively, κ_1 and κ_2 are the concentration parameters, the matrix $\mathbf{A} \in Mat_{2\times 2}$ gives a non-trivial correlation structure, while $\mathcal{Z}$ is the *Partition Function* (a normalization constant) given by:

$$\mathcal{Z} := \int_{[-\pi, +\pi)^2} d\phi d\psi \times$$

$$\exp[\kappa_1 \cos(\phi - \mu) + \kappa_2 \cos(\psi - \nu) + (\cos(\phi - \mu), \sin(\phi - \mu))\mathbf{A}(\cos(\psi - \nu), \sin(\psi - \nu))^T].$$

The FBVM seems to be over-parametrized for being the "Toroidal Counterpart" of the BVN. In fact, the FBVM depends on eight parameters, while the BVN possesses only five parameters. This situation becomes even clearer in the high-concentration limit ($\kappa \to +\infty$)

[87]. For this reason, several submodels have been proposed. Four commonly used variants of the bivariate von Mises distribution have been originally proposed by Mardia [81] and then revisited by Rivest [99] and also by Singh-Hnidzo-Demchuk [110]. These variant are models with a reduced number of parameters and are derived by setting the off-diagonal elements of A to be zero. The following models are submodels of the FBVM which have been discussed in the literature:

- the *Cosine Model with Positive Interaction* [87] has its probability density function as

$$f(\phi,\psi) = \frac{1}{\mathcal{Z}_{c^+}}(\kappa_1,\kappa_2,\kappa_3)\ \exp[\kappa_1\cos(\phi-\mu)+\kappa_2\cos(\psi-\nu)-\kappa_3\cos(\phi-\mu-\psi+\nu)],$$

  where μ and ν are the means for ϕ and ψ, κ_1 and κ_2 their concentration, and κ_3 is related to their correlation, while $\mathcal{Z}_{c^+}$ is the normalization constant and $A = [-\kappa_3, 0; 0, -\kappa_3]$;

- the *Cosine Model with Negative Interaction* [87] has its probability density function as

$$f(\phi,\psi) = \frac{1}{\mathcal{Z}_{c^-}(\kappa_1,\kappa_2,\kappa_3)}\exp[\kappa_1\cos(\phi-\mu)+\kappa_2\cos(\psi-\nu)-\kappa_3\cos(\phi-\mu+\psi-\nu)],$$

  where μ and ν are the means for ϕ and ψ, κ_1 and κ_2 their concentration, and κ_3 is related to their correlation, while $\mathcal{Z}_{c^-}$ is the normalization constant and $A = [-\kappa_3, 0; 0, \kappa_3]$;

- the *Sine Model* [110] has its probability density function as

$$f(\phi,\psi) = \frac{1}{\mathcal{Z}_s(\kappa_1,\kappa_2,\kappa_3)}\ \exp[\kappa_1\cos(\phi-\mu)+\kappa_2\cos(\psi-\nu)+\kappa_3\sin(\phi-\mu)\sin(\psi-\nu)],$$

  where μ and ν are the means for ϕ and ψ, κ_1 and κ_2 their concentration, and κ_3 is related to their correlation, while $\mathcal{Z}_s$ is the normalization constant and $A = [0, 0; 0, -\kappa_3]$;

- the *Hybrid Model* [69] has its probability density function as

$$f(\phi,\psi) = \frac{1}{\mathcal{Z}_h(\kappa_1,\kappa_2,\kappa_3,\epsilon)}\times$$

$$\exp\{\kappa_1\cos(\phi)+\kappa_2\cos(\psi)+\epsilon[(\cosh\gamma-1)\cos\phi\cos\psi+\sinh\gamma\sin\phi\sin\psi]\},$$

  where ϵ is a tuning parameter (often $\epsilon = 1$), μ and ν are the means for ϕ and ψ,

κ_1 and κ_2 their concentration, and κ_3 is related to their correlation, while $\mathcal{Z}_h$ is the normalization constant and $A = [a_{11}, a_{12}; a_{21}, a_{22}]$ satisfying:

$$\begin{cases} a_{11}\cos\mu\cos\nu - a_{12}\cos\mu\sin\nu - a_{21}\sin\mu\cos\nu + a_{22}\sin\mu\sin\nu = \epsilon(\cosh\gamma - 1) \\ a_{11}\sin\mu\sin\nu + a_{12}\sin\mu\cos\nu + a_{21}\cos\mu\sin\nu + a_{22}\cos\mu\cos\nu = \sinh\gamma \end{cases}$$

for some quadruplet of entries $a_{11}, a_{12}, a_{21}, a_{22}$.

The use of these distributions have pros and cons. The pros: the *von Mises* distributions resemble the multivariate normal in the case of high-concentration limit, they are closed with respect to conditioning, it is relatively easy to give multimodality conditions, and the parameters have easy interpretability, even when they do not exactly match the ones of the *Bivariate Normal Distribution* (BVN). The cons: the family is not closed under marginalization (but in the case of high-concentration limit), the estimation and test of hypothesis are not trivial and require numerical methods. For example, the MLEs cannot be computed explicitly and requires optimization algorithms for their determination. To overcome this problem, more advanced procedures, like the pseudo-likelihood estimators, have been suggested [83].

It is a common belief that the geometry of the torus implies that there is not a completely natural counterpart of the BVN on the torus (see, for example, the discussion in [69]). Therefore, despite *von Mises* models have been proven to be successful, there has not been a definite answer yet to which of the models proposed is the “best” candidate to represent the *Toroidal Counterpart* of the BVN.

In this section, we aim at giving an alternative to the *von Mises* type distributions both in univariate and multivariate case, which maintain the good properties of the *von Mises* distributions, like the asymptotic normality, but also possess some extra properties, such as the closure under marginalization or the simplicity in the estimation and hypothesis testing that the MVN possesses in the Euclidean case, but the *von Mises* distributions do not.

The candidate distribution that we propose is called *Inverse Stereographic Normal Distribution on the Torus* (ISND). The name comes from the fact that the *Stereographic Projection* of this distribution is the BVN on $\mathbb{R}^2$ (see Section 4.5.2). We start by giving the definition of the ISND on the circle $\mathbb{S}^1$.

Definition 4.5.1. *We say that a random variable* Θ *has an* Inverse Stereographic Normal

Distribution*, denote by* $\mathcal{ISN}(\mu,\sigma^2)$*, if its probability density function is given by*

$$f_\Theta(\theta) := \sqrt{\frac{1}{2\pi}}\frac{1}{(1+\cos\theta)\sigma}e^{-\frac{1}{2\sigma^2}\left(\frac{\sin\theta}{1+\cos\theta}-\mu\right)^2},$$

for some $\mu \in \mathbb{R}$*,* $\sigma > 0$ *and for any* $\theta \in [-\pi, +\pi)$*.*

In a similar manner, we can define the *Multivariate Inverse Stereographic Normal Distribution*, as follows.

Definition 4.5.2. *We say that a random variable* $\mathbf{\Theta} := \{\Theta_1, \ldots, \Theta_n\}$ *has a* Multivariate Inverse Stereographic Normal Distribution*, denoted by* $\mathcal{MISN}(\mu, \mathbf{\Sigma})$*, if its probability density function is given by*

$$f_{\mathbf{\Theta}}(\theta) :=$$

$$\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi|\Sigma|}(1+\cos\theta_i)}\right)e^{-\frac{1}{2}\left(\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\ldots,\frac{\sin\theta_n}{1+\cos\theta_n}-\mu_n\right)^T\Sigma^{-1}\left(\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\ldots,\frac{\sin\theta_n}{1+\cos\theta_n}-\mu_n\right)},$$

for some $\mu := (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n$*,* $\Sigma \in Sym^+_{n\times n}$ *and for any* $\theta = (\theta_1, \ldots, \theta_n) \in [-\pi, +\pi)^n$*.*

Remark 4.5.3. *In Section 4.5.2, we will see that this type of construction can be used for all probability distributions on the Euclidean space.*

As we will see in the subsequent subsections, there are several advantages of this approach, listed below not necessarily in order of importance:

- The *Stereographic Projection* suggests a natural way to construct distributions on manifolds, by transforming distributions on the Euclidean Space (see Section 4.5.2);
- The number of parameters of the ISND match the number of parameters of the BVN, without imposing any further constraint (see Section 4.5.2);
- The *Bivariate Von Mises Distribution* and ISND resemble each other in the case of high-concentration limit for certain ranges of parameters and, moreover, the ISND approximates the BVN in the case of high concentration limit (see Section 4.5.3);
- There is a natural interpretation for the parameters of the ISND, relating to the parameters of the BVN (see Section 4.5.4);
- The *Stereographic Projection* suggests a more geometric counterpart of the *Euclidean Moments* which differs from the *Circular Moments* (see Section 4.5.4);
- The definition of moments using the *Stereographic Projection* helps address some problems of interpretation of the parameters of directional variables (see Remark 4.5.45 in Section 4.5.6.1);

- The estimation problems for the ISND do not need numerical methods, since the estimates of the parameters are the transformed estimates of the parameters of the BVN (see Section 4.5.6.1);
- The *Stereographic Projection* gives the possibility of transferring the test statistics from the Euclidean space to Manifolds and this simplifies the problems of Hypothesis Testing, Confidence Intervals and Goodness of Fit (see Section 4.5.6.1).

The rest of this section is organized as follows.

In Subsection 4.5.2, we describe the construction of general inverse stereographic projected distributions and present the main example of the ISND. We give results about the marginals of the ISND which are still ISND (Theorem 4.5.13), and conditionals which are ISND as well (Theorem 4.5.16). We also present some results on bimodality conditions for this family of distributions (Theorem 4.5.17 and Theorem 4.5.19).

In Subsection 4.5.3, we discuss the comparison of the ISND and the BVM. In particular, we prove an approximation result in the case of high-concentration limit which connects BVN, ISDN and BVM models for some subset of the parameter space.

In Subsection 4.5.4, we introduce inverse stereographic moments and moment generating functions which lead naturally to a corresponding central limit theorem on the torus. We compare the moments with the classical circular moments and with the moments of Euclidean random variables.

In Subsection 4.5.6, we pass to statistical inference. We first discuss point estimation, confidence intervals and hypothesis testing for the model parameters (see Subsection 4.5.6.1). Then, we discuss sampling methods (see Subsection 4.5.7). All these constructions follow the lines of Euclidean cases.

In Subsection 4.5.8, we present some numerical examples and applications. In Subsubsection 4.5.8.1, we produce some plots of the ISND for some choices of the parameters, both in the unimodal and the multimodal cases. In Subsubsection 4.5.8.2, we compare numerically the BVM and the ISND, using Theorem 4.5.21 for the corresponding ranges of the parameters. We conclude with Subsection 4.5.8.4, where we give an application to a problem in molecular biology.

Notation and Preliminaries

In the following, we will use the notation $Mat_{n\times n}$ to denote the set of square matrices of dimension n and $Sym^+_{n\times n}$ to denote positive symmetric matrices of dimension n.

4.5.2 Inverse Stereographic Projected Distributions on the Torus

In this subsection, we introduce the Inverse Stereographic Projected Distributions and then give some properties of the model. As mentioned in the introduction, here we describe the construction of general inverse stereographic projected distributions and then present the main example of the ISND. We present results about the marginals of ISND which are still ISND (Theorem 4.5.13), and conditionals which are ISND as well (Theorem 4.5.16). We finally present some results on bimodality conditions for this family of distributions.

4.5.2.1 Inverse Stereographic Projected Probability Distributions

In this subsubsection, we describe how to stereographic project a probability distribution defined on $\mathbb{S}^{n-1}$ or $\mathbb{T}^n$ into the Euclidean Space $\mathbb{R}^n$, and how to construct the inverse projected distribution. Let us start with the one dimensional case $n = 1$.

Theorem 4.5.4. *Suppose $f_X(x)$ is a pdf on $\mathbb{R}$, and we consider the* Stereographic Projection *in angular coordinates*

$$\theta \mapsto x := \frac{\sin\theta}{1+\cos\theta},$$

with $\theta \in [-\pi, +\pi)$. Then,

$$f_\Theta(\theta) = \frac{1}{1+\cos\theta} f_X\left(\frac{\sin\theta}{1+\cos\theta}\right).$$

Vice versa, suppose $f_\Theta(\theta)$ is a pdf on $\mathbb{S}^1$, and we consider the Inverse Stereographic Projection

$$x \mapsto \theta := 2\arctan(x),$$

for $x \in \mathbb{R}$. Then,

$$f_X(x) = \frac{2}{1+x^2} f_\Theta\left(2\arctan(x)\right).$$

Proof. Follows by a simple change of variables formula. □

Remark 4.5.5. *This theorem has an analogous counterpart in higher dimensions, where* *the metric of* $\mathbb{S}^{n-1}$ *is given by*

$$g_{\mathbb{S}^{n-1}}(x) := \frac{4}{(1+x_1^2+\cdots+x_n^2)^2}\left(dx_1^2+\cdots+dx_{n-1}^2\right).$$

A simple consequence of this theorem is a characterization of the *Cauchy Distribution* on the real line by means of the *Stereographic Projection*.

Corollary 4.5.6. *The Cauchy Distribution on* $\mathbb{R}$ *is the unique distribution mapped to the Uniform distribution* $\mathbb{S}^1$ *under* Stereographic Projection.

Proof. It is straightforward to show this. □

Theorem 4.5.4 provides us a formula which works also for the torus $\mathbb{T}^n := \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$.

Theorem 4.5.7. *Suppose* $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ *is a pdf on* $\mathbb{R}^n$, *and we consider the* Stereographic Projection

$$(\theta_1,\ldots\theta_n) \mapsto (x_1,\ldots,x_n) := \left(\frac{\sin(\theta_1)}{1+\cos(\theta_1)},\ldots,\frac{\sin(\theta_n)}{1+\cos(\theta_n)}\right),$$

for $(\theta_1,\ldots,\theta_n) \in \mathbb{T}^n$ *and so with* $\theta_i \in [-\pi,+\pi)$ *for every* $i=1,\ldots,n$. *Then,*

$$f_{\Theta_1,\ldots,\Theta_n}(\theta_1,\ldots,\theta_n)d\theta_1\ldots d\theta_n = f_{X_1,\ldots,X_n}\left(\frac{\sin(\theta_1)}{1+\cos(\theta_1)},\ldots,\frac{\sin(\theta_n)}{1+\cos(\theta_n)}\right)\prod_{i=1}^{n}\frac{d\theta_i}{1+\cos(\theta_i)}.$$

Vice versa, suppose $f_{\Theta_1,\ldots,\Theta_n}(\theta_1,\ldots,\theta_n)$ *is a pdf on* $\mathbb{T}^n$, *and we consider the* Inverse Stereographic Projection

$$(x_1,\ldots,x_n) \mapsto (2\arctan(x_1),\ldots,2\arctan(x_n)),$$

for every $(x_1,\ldots,x_n) \in \mathbb{R}^n$. *Then,*

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)dx_1\ldots dx_n = 2^n\prod_{i=1}^{n}\frac{dx_i}{1+x_i^2}f_{\Theta_1,\ldots,\Theta_n}\left(2\arctan(x_1),\ldots,2\arctan(x_n)\right).$$

Proof. this can be shown by a straightforward calculation. □

Remark 4.5.8. *We give a brief remark about the spherical case. To fix the ideas, we consider the case of dimension* $n=3$. *In this case, the stereographic projection is given by*

$$(\theta,\phi) \mapsto (x,y) := \left(\frac{\cos(\theta)\sin(\phi)}{1-\cos(\phi)},\frac{\sin(\theta)\sin(\phi)}{1-\cos(\phi)}\right),$$

for $\theta \in [-\pi, +\pi)$ *and* $\phi \in [0, +\pi)$. *Therefore, we have*

$$f_{\Theta,\Phi}(\theta,\phi)d\theta d\phi = f_{X,Y}\left(\frac{\cos(\theta)\sin(\phi)}{1-\cos(\phi)}, \frac{\sin(\theta)\sin(\phi)}{1-\cos(\phi)}\right)\frac{\sin(\phi)d\theta d\phi}{(1-\cos(\phi))^2},$$

since the Jacobian of the transformation is $J = \frac{\sin(\phi)}{(1-\cos(\phi))^2}$.

Vice versa, suppose $f_{\Theta,\Phi}(\theta,\phi)$ *is a pdf on* $\mathbb{S}^2$, *and we consider the inverse* Stereographic Projection

$$(x,y) \mapsto (\theta,\phi) = \left(\arctan\left(\frac{y}{x}\right), \arccos\left(\frac{x^2+y^2-1}{x^2+y^2+1}\right)\right),$$

for every $(x,y) \in \mathbb{R}^2$. *Then,*

$$f_{X,Y}(x,y)dxdy = f_{\Theta,\Phi}\left(\arctan\left(\frac{y}{x}\right), \arccos\left(\frac{x^2+y^2-1}{x^2+y^2+1}\right)\right)\frac{2dxdy}{(1+x^2+y^2)\sqrt{x^2+y^2}},$$

by the inverse formula for the Jacobian.

Note that, even if in both the spherical and toroidal cases, the density functions on the manifold and on the Euclidean Space are defined through a Stereographic Projection, *the final density functions are different and do reflect the geometry of the manifold.*

Remark 4.5.9. *The* Stereographic Projection *is a* Conformal Transformation *and so it does not change angles. In our context, this is a very useful property as it means that the* Stereographic Projection *maps* elliptical contours *to* elliptical contours.

4.5.2.2 Inverse Stereographic Projected Normal Distribution

We are now ready to define the *Inverse Stereographic Normal Distribution.*

Definition 4.5.10. *We say that a random variable* Θ *has an* Inverse Stereographic Standard Normal Distribution, *denoted by* $\mathcal{ISN}(0,1)$, *if and only if its density function is given by*

$$f_\Theta(\theta) := \sqrt{\frac{1}{2\pi}}\frac{1}{1+\cos\theta}e^{-\frac{1}{2}\left(\frac{\sin\theta}{1+\cos\theta}\right)^2},$$

for $\theta \in [-\pi, \pi)$.

We can then introduce some parameters to the model and give a general definition of the ISND.

Definition 4.5.11. *We say that a random variable* Θ *has an* Inverse Stereographic Normal Distribution*, denoted by* $\mathcal{ISN}(\mu, \sigma^2)$*, if and only if its density function is given by the following formula:*

$$f_\Theta(\theta) := \sqrt{\frac{1}{2\pi}} \frac{1}{(1+\cos\theta)\sigma} e^{-\frac{1}{2\sigma^2}\left(\frac{\sin\theta}{1+\cos\theta}-\mu\right)^2},$$

for some $\mu \in \mathbb{R}$*,* $\sigma > 0$ *and for any* $\theta \in [-\pi, +\pi)$*.*

In a similar way, we can define the *Multivariate Inverse Stereographic Normal Distribution.*

Definition 4.5.12. *We say that a random variable* $\mathbf{\Theta} := \{\Theta_1, \ldots, \Theta_n\}$ *has a* Multivariate Inverse Stereographic Normal Distribution*, denoted by* $\mathcal{MISN}(\mu, \mathbf{\Sigma})$*, if and only if its density function is given by*

$$f_{\mathbf{\Theta}}(\theta) :=$$

$$\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi|\Sigma|}(1+\cos\theta_i)}\right) e^{-\frac{1}{2}\left(\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\ldots,\frac{\sin\theta_n}{1+\cos\theta_n}-\mu_n\right)^T \Sigma^{-1}\left(\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\ldots,\frac{\sin\theta_n}{1+\cos\theta_n}-\mu_n\right)},$$

for some $\mu := (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$*,* $\Sigma \in Sym^+_{n\times n}$ *and for any* $\theta = (\theta_1, \ldots, \theta_n) \in [-\pi, +\pi)^n$*.*

In strong contrast with the *Multivariate Von Mises Distribution*, the *Inverse Stereographic Normal Distribution* is closed under marginalization.

Theorem 4.5.13. *Suppose that a random variable* $\mathbf{\Theta} := \{\Theta_1, \ldots, \Theta_n\} \sim \mathcal{MISN}(\mu, \mathbf{\Sigma})$*. Then, for every* $i = 1, \ldots, n$*, we have* $\Theta_i \sim \mathcal{ISN}(\mu_i, \Sigma_{ii})$*.*

Proof. We give the proof just in the case $n = 2$, which is more interesting for the applications that we have in mind, but the proof for the general case follows in a similar way.

Suppose a random variable $\mathbf{\Theta} := \{\Theta_1, \ldots, \Theta_n\} \sim \mathcal{MISN}(\mu, \mathbf{\Sigma})$. Its pdf is given by

$$f_{\mathbf{\Theta}}(\theta) :=$$

$$\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi|\Sigma|}(1+\cos\theta_i)}\right) e^{-\frac{1}{2}\left(\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\ldots,\frac{\sin\theta_n}{1+\cos\theta_n}-\mu_n\right)^T \Sigma^{-1}\left(\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\ldots,\frac{\sin\theta_n}{1+\cos\theta_n}-\mu_n\right)},$$

for some $\mu := (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$, some $\Sigma \in Sym^+_{n\times n}$ and for any $\theta = (\theta_1, \ldots, \theta_n) \in [-\pi, +\pi)^n$. In the case $n = 2$, we have:

$$f_{\Theta_1,\Theta_2}(\theta_1, \theta_2) := \frac{1}{2\pi|\Sigma|} \frac{1}{(1+\cos\theta_1)} \frac{1}{(1+\cos\theta_2)}$$

$$\times e^{-\frac{1}{2}\left[\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\frac{\sin\theta_2}{1+\cos\theta_2}-\mu_2\right]^T\Sigma^{-1}\left[\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\frac{\sin\theta_2}{1+\cos\theta_2}-\mu_2\right]}.$$

Therefore,

$$f_{\Theta_1}(\theta_1) = \int_{-\pi}^{+\pi} d\theta_2 f_{\Theta_1,\Theta_2}(\theta_1,\theta_2) = \int_{-\pi}^{+\pi} d\theta_2 \frac{1}{2\pi|\Sigma|}\frac{1}{(1+\cos\theta_1)}\frac{1}{(1+\cos\theta_2)}\times$$

$$\times e^{-\frac{1}{2}\left[\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\frac{\sin\theta_2}{1+\cos\theta_2}-\mu_2\right]^T\Sigma^{-1}\left[\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1,\frac{\sin\theta_2}{1+\cos\theta_2}-\mu_2\right]}.$$

Now, define $B := \Sigma^{-1}$ and so $B_{ij} = (\Sigma^{-1})_{ij}$ for $j = 1, 2$ and $2b = b_{12} + b_{21}$. This implies that

$$f_{\Theta_1}(\theta_1) = \frac{1}{2\pi|\Sigma|}\frac{e^{-\frac{b_{11}}{2}\left\{\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1\right\}^2}}{(1+\cos\theta_1)}$$

$$\times \int_{-\pi}^{+\pi}\frac{d\theta_2}{(1+\cos\theta_2)}e^{-\frac{b_{22}}{2}\left\{\frac{\sin\theta_2}{1+\cos\theta_2}-\mu_2\right\}^2}e^{-\frac{2b}{2}\left\{\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1\right\}\left\{\frac{\sin\theta_2}{1+\cos\theta_2}-\mu_2\right\}}.$$

Now, we do the following change of variables:

$$z := \frac{\sin\theta_2}{1+\cos\theta_2} - \mu_2, \quad dz = \frac{d\theta_2}{1+\cos\theta_2}.$$

Then, we obtain:

$$f_{\Theta_1}(\theta_1) = \frac{1}{2\pi|\Sigma|}\frac{e^{-\frac{b_{11}}{2}\left\{\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1\right\}^2}}{(1+\cos\theta_1)}\int_{-\infty}^{+\infty} dz e^{-\frac{1}{2}\left[2b\alpha(\theta_1)z+b_{22}z^2\right]},$$

with $\alpha(\theta_1) := \frac{\sin\theta_1}{1+\cos\theta_1} - \mu_1$. Upon completing the square, we get

$$f_{\Theta_1}(\theta_1) = \frac{1}{2\pi|\Sigma|}\frac{e^{-\frac{b_{11}}{2}\left\{\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1\right\}^2}e^{+\frac{b^2}{2b_{22}}\left\{\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1\right\}^2}}{(1+\cos\theta_1)}\int_{-\infty}^{+\infty} dz e^{-\frac{1}{2}\left[b_{22}z+\frac{b\alpha(\theta_1)}{\sqrt{b_{22}}}\right]^2}.$$

By performing another change of variables

$$x := b_{22}z + \frac{b\alpha(\theta_1)}{\sqrt{b_{22}}}, \quad dx = b_{22}dz,$$

we get

$$f_{\Theta_1}(\theta_1) = \frac{1}{2\pi|\Sigma|b_{22}}\frac{e^{-\frac{b_{11}}{2}\left\{\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1\right\}^2}e^{+\frac{b^2}{2b_{22}}\left\{\frac{\sin\theta_1}{1+\cos\theta_1}-\mu_1\right\}^2}}{(1+\cos\theta_1)}\int_{-\infty}^{+\infty} dx e^{-\frac{1}{2}x^2}$$

$$= \frac{b_{22}b_{11} - b^2}{\sqrt{2\pi} b_{22}} \frac{e^{-\frac{1}{2}\frac{b_{11}b_{22}-b^2}{b_{22}}\left\{\frac{\sin\theta_1}{1+\cos\theta_1} - \mu_1\right\}^2}}{(1+\cos\theta_1)} = \frac{1}{\sqrt{2\pi}\Sigma_{11}} \frac{e^{-\frac{1}{2\Sigma_{11}}\left\{\frac{\sin\theta_1}{1+\cos\theta_1} - \mu_1\right\}^2}}{(1+\cos\theta_1)}.$$

Since

$$f_{\Theta_1}(\theta_1) = \frac{1}{\sqrt{2\pi}\Sigma_{11}} \frac{e^{-\frac{1}{2\Sigma_{11}}\left\{\frac{\sin\theta_1}{1+\cos\theta_1} - \mu_1\right\}^2}}{(1+\cos\theta_1)},$$

then $\Theta_1 \sim ISN(\mu_1, \Sigma_{11})$. Thus, the family of *Multivariate Inverse Stereographic Normal Distributions* is closed under marginalization, as required. □

Remark 4.5.14. *Note that, left as is, the $\mathcal{MISN}(\mu, \mathbf{\Sigma})$ might not seem completely satisfactory for the following reason. Suppose for simplicity that we consider the case of $\mathbb{S}^1$. We have $P_\Theta(\Theta \in [\alpha + \theta_0, \beta + \theta_0])$ which depends on θ_0, and thus it depends on the choice of the north pole in the* Stereographic Projection*. This does not seem a desirable property, but the choice of the North Pole corresponds, in some sense, to the choice of the point at infinity in the real line. Also in the real line, we make the arbitrary choice of placing the origin somewhere and we somehow "break the symmetry" of the real line, as we do here for the circle.*

Remark 4.5.15. *All this suggests us to consider a natural measure on the manifold, at least when the probability distribution is obtained through* Stereographic Projection*. This measure is the pull-back measure*

$$P^* dx = \frac{d\theta}{1+\cos\theta},$$

where P is the Stereographic Projection*. This suggests a different way to compute moments and the moment generating function. We discuss this in more detail later in Section 4.5.4.*

Similarly, the *Inverse Stereographic Normal Distribution* is also closed under conditionals.

Theorem 4.5.16. *Suppose a random variable $\mathbf{\Theta} := \{\Theta_1, \ldots, \Theta_n\} \sim \mathcal{MISN}(\mu, \mathbf{\Sigma})$ and we have the partition $\Theta = (\Theta_1, \Theta_2)$, $\mu = (\mu_1, \mu_2)$,*

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where Θ_1 and μ_1 are $k \times 1$ vectors and $\Sigma_{11} \in Mat_{k\times k}$, and Σ_{22}^- is the generalized inverse *of Σ_{22}. Then, the conditional distribution of Θ_1, given Θ_2 is given by*

$$\Theta_1 \mid \Theta_2 \sim \mathcal{MISN}(\mu_1 - \Sigma_{12}\Sigma_{22}^{-}(\Theta_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}).$$

Proof. The proof follows the same lines as in the proof of Theorem 4.5.13 and the techniques used in the Euclidean case (see for example Theorem 1.2.11 page 12 in [90]). □

4.5.2.3 Unimodality and Multimodality Conditions

As can be seen from some simulations (see, for example, Subsection 4.5.8.1), ISND can be multimodal and so it is important to determine the conditions on the parameters which ensure that the pdf is unimodal and the ones that produce a multimodal distribution.

We now give necessary and sufficient conditions for the ISND to be unimodal.

Consider the density function of a ISND with parameters $\mu = 0$ and $\sigma^2 > 0$, given by

$$f_\Theta(\theta) := \sqrt{\frac{1}{2\pi}} \frac{1}{1+\cos\theta} e^{-\frac{1}{2}\left(\frac{sin\theta}{1+\cos\theta}\right)^2}.$$

Note that for $\theta = 0$ the density has a removable singularity and it can be extended to a smooth function by choosing $f_\Theta(\pi) = 0$. This makes $\theta = 0$ a critical point of the density and hence the global minimum. Since $\mathbb{S}^1$ is compact and $f_\Theta(\theta)$ is not constant, it admits a maximum and a minimum distinct from the maximum. To ensure unimodality, we need to find μ and σ^2 such that the derivative of $f_\Theta(\theta)$ admits only one further zero apart from $\theta = \pi$.

Then, we start by computing the first derivative of the pdf $f_\Theta(\theta)$ as

$$\frac{d}{d\theta} f_\Theta(\theta) = \sqrt{\frac{1}{2\pi\sigma^2}} \frac{1}{(1+\cos\theta)^2\sigma^2} e^{-\frac{1}{2\sigma^2}\left(\frac{\sin\theta}{1+\cos\theta}-\mu\right)^2} \times \left\{\sigma^2 \sin\theta - \left(\frac{\sin\theta}{1+\cos\theta} - \mu\right)\right\}.$$

Therefore, we need to find conditions for which

$$\sigma^2 \sin\theta - \left(\frac{\sin\theta}{1+\cos\theta} - \mu\right) = 0$$

only once for $\theta \in [-\pi, +\pi)$.

Note that in the simplified case, when $\mu = 0$, we need to solve

$$\sigma^2 \sin\theta - \frac{\sin\theta}{1+\cos\theta} = 0,$$

which has the solutions $\sin\theta = 0$ so that $\theta = 0$ and $\theta = \pi$ (already counted) and $\cos\theta = \frac{1}{\sigma^2} - 1$, which has solutions just when $-1 \leq \frac{1}{\sigma^2} - 1 \leq 1$, which implies

$$\sigma^2 \geq \frac{1}{2}.$$

Therefore, we have the following theorem.

Theorem 4.5.17. *Consider the* Inverse Stereographic Normal Distribution $f_\Theta(\theta|\mu, \sigma^2)$ *with parameters* $\mu = 0$ *and* $\sigma^2 > 0$*. Then,* $f_\Theta(\theta|\mu, \sigma^2)$ *is unimodal if and only if* $\sigma^2 < \frac{1}{2}$*.*

The interpretation here is that when the mass of the density is too spread out on the real line, when sent to the circle, it tends to accumulate on itself and produce more than one peak.

The general case is much more complicated and does not have a solution that is easily computable, because it needs the solution formula for general polynomial equations of third order. This is what we consider next.

We need to solve the equation

$$\sigma^2 \sin\theta - \left(\frac{\sin\theta}{1+\cos\theta} - \mu\right) = 0$$

for $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. We actually need to prove that it has only a single solution for $\theta \in [-\pi, +\pi)$. We have already discussed the case $\theta = \pi$, and so we can multiply the equality by $1 + \cos\theta$ (which is non-zero) to get:

$$\sigma^2 \sin\theta\cos\theta + \sigma^2 \sin\theta - \sin\theta + \mu(1+\cos\theta) = 0.$$

Therefore,

$$\cos\theta(\sigma^2 \sin\theta + \mu) = \sin\theta - \mu - \sigma^2 \sin\theta$$

and hence

$$\left(1 - \sin^2\theta\right)\left(\sigma^2 \sin\theta + \mu\right)^2 = \left(\sin\theta - \mu - \sigma^2 \sin\theta\right)^2.$$

By expanding the terms, we get

$$\sigma^4 \sin^2\theta + 2\mu\sigma^2 \sin\theta + \mu^2 - \sigma^4 \sin^4\theta - 2\mu\sigma^2 \sin^3\theta - \mu^2 \sin^2\theta$$

$$= \sin^2\theta + \mu^2 + \sigma^4 \sin^2\theta - 2\mu \sin\theta - 2\sigma^2 \sin^2\theta + 2\mu\sigma^2 \sin\theta.$$

Upon simplifications, we get

$$-\sigma^4 \sin^4\theta - 2\mu\sigma^2 \sin^3\theta + \sin^2\theta \left(-\mu^2 - 1 + 2\sigma^2\right) + \sin\theta(2\mu) = 0,$$

and so

$$\sigma^4 \sin^4\theta + 2\mu\sigma^2 \sin^3\theta + \sin^2\theta \left(\mu^2 + 1 - 2\sigma^2\right) - 2\mu \sin\theta = 0$$

which factors to

$$\sin\theta \left(\sigma^4 \sin^3\theta + 2\mu\sigma^2 \sin^2\theta + \sin\theta \left(\mu^2 + 1 - 2\sigma^2\right) - 2\mu\right) = 0.$$

Note that $\theta = 0$ implies $\mu = 0$ which we have already treated before. The case $\theta = \pi$ has already been discussed too. So, we reduce our problem to solving

$$\sigma^4 \sin^3\theta + 2\mu\sigma^2 \sin^2\theta + \sin\theta \left(\mu^2 + 1 - 2\sigma^2\right) - 2\mu = 0.$$

The precise three solutions of this system are not of extreme importance and the computations are in fact pretty involved. In any case, they can be computed pretty easily using some symbolic software like Maple or Wolfram Alpha. We got, for $\sigma^2 \neq 0$, the following

$$x_1 = \sin\theta_1 = -\frac{2\mu}{3\sigma^2} - \frac{1}{81\sqrt[3]{2}\sigma^6}((-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)$$

$$+\sqrt{((-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)^2 + 4(2187\sigma^8 - 4374\sigma^{10} - 729\sigma^8\mu^2)^3))^{1/3})}$$

$$+(\sqrt{2}(2187\sigma^8 - 4374\sigma^{10} - 729\sigma^8\mu^2))/$$

$$(81\sigma^6(-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu$$

$$+\sqrt{((-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)^2 + 4(2187\sigma^8 - 4374\sigma^{10} - 729\sigma^8\mu^2)^3))^{1/3}),}$$

$$x_2 = \sin\theta_2 = -\frac{2\mu}{3\sigma^2} - \frac{1}{162\sqrt[3]{2}\sigma^6}(1 - i\sqrt{3})((-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)$$

$$+\sqrt{((-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)^2 + 4(2187\sigma^8 - 4374\sigma^{10} - 729\sigma^8\mu^2)^3)})^{1/3})$$

$$-((1 + i\sqrt{3})(2187\sigma^8 - 4374\sigma^{10} - 729\sigma^8\mu^2))/$$

$$(81 * 2^{2/3}\sigma^6(-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)$$

$$+\sqrt{((-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)^2 + 4(2187\sigma^8 - 4374\sigma^{10} - 729\sigma^8\mu^2)^3)})^{1/3})$$

and

$$x_3 = \sin\theta_3 = -\frac{2\mu}{3\sigma^2} - \frac{1}{162\sqrt[3]{2}\sigma^6}(1 + i\sqrt{3})((-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)$$

$$+\sqrt{((-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)^2 + 4(2187\sigma^8 - 4374\sigma^{10} - 729\sigma^8\mu^2)^3)})^{1/3})$$

$$-((1 - i\sqrt{3})(2187\sigma^8 - 4374\sigma^{10} - 729\sigma^8\mu^2))/$$

$$81 * 2^{2/3}\sigma^6(-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)$$

$$+\sqrt{((-354294\sigma^{14}\mu - 39366\sigma^{12}\mu^3 - 354294\sigma^{12}\mu)^2 + 4(2187\sigma^8 - 4374\sigma^{10} - 729\sigma^8\mu^2)^3)})^{1/3}).$$

What we actually want is to find conditions on μ and σ^2 so that x_2, x_3 are really complex conjugate. If we manage to do so, we will just have that $x_1 = x_1(\mu, \sigma^2)$ is the only possible other critical point for $f_\Theta(\theta)$ apart from $\theta = \pi$.

For this, we need to examine the *Discriminant of the Polynomial* given by

$$\sigma^4 \sin^3\theta + 2\mu\sigma^2 \sin^2\theta + \sin\theta\left(\mu^2 + 1 - 2\sigma^2\right) - 2\mu = 0.$$

We recall the following lemma.

Lemma 4.5.18. *Consider the general cubic equation*

$$ax^3 + bx^2 + cx + d = 0$$

with $a \neq 0$ and $a, b, c, d \in \mathbb{R}$. This cubic equation, with real coefficients, has at least one real solution x_1. Moreover, consider the Discriminant of the Polynomial*:*

$$ax^3 + bx^2 + cx + d = 0$$

given by

$$\Delta := 18abcd - 4b^3d + b^2c^2 - 4ac^3 - 27a^2d^2.$$

Then:

- *If $\Delta > 0$, the equation has three distinct real roots;*
- *If $\Delta = 0$, the equation has multiple roots and all the roots are real;*
- *If $\Delta < 0$, the equation has one real root and two complex conjugate roots.*

We want to make use of this lemma and compute the *Discriminant of the Polynomial*:

$$\sigma^4 \sin^3\theta + 2\mu\sigma^2 \sin^2\theta + \sin\theta\left(\mu^2 + 1 - 2\sigma^2\right) - 2\mu = 0.$$

Note that this polynomial is cubic in the variable $x = \sin\theta$. In the notation of Lemma 4.5.18, we have

$$a = \sigma^4, \quad b = 2\mu\sigma^2, \quad c = \mu^2 + 1 - 2\sigma^2, \quad d = -2\mu.$$

Therefore, the discriminant $\Delta = \Delta(\mu, \sigma^2)$ in this case becomes

$$\Delta(\mu, \sigma^2) = -72\mu\sigma^2(\mu\sigma^4)(\mu^2 + 1 - 2\sigma^2) + 8\mu(2\mu\sigma^2)^3 - 4\sigma^4(\mu^2 + 1 - 2\sigma^2)^3 - 108\sigma^8\mu^2$$

$$= -72\mu^4\sigma^8 - 72\mu\sigma^4 + 144\mu\sigma^6 + 64\mu^4\sigma^6 - 4\sigma^4$$

$$\times(\mu^6 - 6\mu^4\sigma^2 + 3\mu^4 + 12\mu^2\sigma^4 - 12\mu^2\sigma^2 + 3\mu^2 - 8\sigma^6 + 12\sigma^4 - 6\sigma^2 + 1) - 108\sigma^8\mu^2.$$

It seems complicated to find more explicit analytic conditions for which $\Delta(\mu, \sigma^2) < 0$, and so we leave this criterion as is. To verify the unimodality, it is enough to plug in the values of μ and σ^2 and verify that one obtains a negative value which would ensure unimodality. We have thus just proved the following theorem.

Theorem 4.5.19. *Consider the* Inverse Stereographic Normal Distribution $f_\Theta(\theta|\mu, \sigma^2)$ *with parameters* $\mu \neq 0$ *and* $\sigma^2 > 0$*. Then,* $f_\Theta(\theta|\mu, \sigma^2)$ *is unimodal if and only if* $\Delta(\mu, \sigma^2) < 0$*, where* $\Delta(\mu, \sigma^2)$ *is given by*

$$\Delta(\mu, \sigma^2) := -72\mu^4\sigma^8 - 72\mu\sigma^4 + 144\mu\sigma^6 + 64\mu^4\sigma^6 - 4\sigma^4$$

$$\times(\mu^6 - 6\mu^4\sigma^2 + 3\mu^4 + 12\mu^2\sigma^4 - 12\mu^2\sigma^2 + 3\mu^2 - 8\sigma^6 + 12\sigma^4 - 6\sigma^2 + 1) - 108\sigma^8\mu^2.$$

Remark 4.5.20. *Note that for* $\mu = 0$*, the discriminant* $\Delta(\mu, \sigma^2) = -4(-8\sigma^6 + 12\sigma^4 - 6\sigma^2 + 1)$ *is negative if and only if* $0 \leq \sigma^2 \leq \frac{1}{2}$ *and so this is the case of unimodality. This is in agreement with what we found above.*

4.5.3 Inverse Stereographic Normal vs Von Mises Models

In this subsection, we compare the *Inverse Stereographic Normal Distribution* with the *Von Mises Distribution*. A classical argument to promote the use of the *Von Mises Distribution* as a natural circular counterpart of the *Normal Distribution* is due to the fact that, in the case of high-concentration limit, the two distributions resemble each others

$$f_{VM}(\theta) \propto e^{\kappa_1 \cos\theta} \simeq e^{-\kappa_1\theta^2/2},$$

for $0 < \theta \ll 1$ or for $\kappa \gg 1$. Note that this approximation does not uniquely identify the *Von Mises Distribution*, and in fact it is valid for several other distributions as well. In particular, it is valid also for the *Inverse Stereographic Standard Normal Distribution.*

We note that, for most of the choices of parameters, the Bivariate von Mises and the Inverse Stereographic Normal are not close to each other, but they are in some particular cases. We analyse these cases and give precise asymptotic bounds.

Theorem 4.5.21. *Consider the* Full Bivariate Von Mises Distribution*:*

$$f_{VM}(\phi, \psi)$$

$$\propto \exp[\kappa_1 \cos(\phi - \mu) + \kappa_2 \cos(\psi - \nu) + (\cos(\phi - \mu), \sin(\phi - \mu))\mathbf{A}(\cos(\psi - \nu), \sin(\psi - \nu))^T],$$

defined for $\phi \in [-\pi, +\pi)$ and $\psi \in [-\pi, +\pi)$. Here, $\mu \in [-\pi, +\pi)$ and $\nu \in [-\pi, +\pi)$ represent the marginal mean values, κ_1 and κ_2 are the concentration parameters, and the matrix $\mathbf{A} \in Mat_{2\times 2}$ is related to their correlation. Further, consider the Bivariate Inverse Stereographic Normal Distribution*:*

$$f_{SN}(\phi, \psi)$$

$$:= \frac{1}{2\pi|\Sigma|} \frac{1}{1+\cos\phi,} \frac{1}{1+\cos\psi} e^{-\frac{1}{2}\left(\frac{\sin\phi,}{1+\cos\phi,} - \mu_1, \frac{\sin\psi}{1+\cos\psi} - \mu_2\right)^T \Sigma^{-1}\left(\frac{\sin\phi,}{1+\cos\phi,} - \mu_1, \frac{\sin\psi}{1+\cos\psi} - \mu_2\right)},$$

defined for $\phi \in [-\pi, +\pi)$ and $\psi \in [-\pi, +\pi)$. Here, $\mu := (\mu_1, \mu_2) \in \mathbb{R}^2$, $\Sigma \in Sym^+_{2\times 2}$. Moreover, suppose $\mu = \nu = \mu_1 = \mu_2 = 0$, $b = a_{12} = a_{21} = 0$ and $b_{11} = \kappa_{11} + (a_{11} + a_{22})$, $b_{22} = \kappa_{22} + (a_{11} + a_{22})$. Then, the following holds:

$$\|f_{ISN}(\phi, \psi) - f_{VM}(\phi, \psi)\| \le C\left(|\phi|^3 + |\psi|^3\right),$$

for $(\phi, \psi) \in B_\epsilon(0, 0)$ with $0 < \epsilon \ll 1$.

Remark 4.5.22. *Note that the result can be restated as an high-concentration limit approximation result.*

Remark 4.5.23. *This theorem reveals that the choice of the* Von Mises Distribution *as the spherical counterpart of the* Normal Distribution *is just one of the possible choices and not the only choice, since there are several distributions with the same asymptotic behaviour in the case of high-concentration limit.*

Proof. Consider the *Bivariate von Mises Distribution*, upon expanding the trigonometric functions and taking log of the pdf, we get:

$$\log(f_{VM}(\phi, \psi)) \propto \kappa_1[cos\phi \cos\mu_1 + \sin\phi \sin\mu_1] + \kappa_2[cos\psi \cos\mu_2 + \sin\psi \sin\mu_2]$$

$$+a_{11}[cos\phi \cos\mu_1 + \sin\phi \sin\mu_1][cos\psi \cos\mu_2 + \sin\psi \sin\mu_2]$$

$$+a_{12}[cos\phi \cos\mu_1 + \sin\phi \sin\mu_1][\sin\psi \cos\mu_2 - \sin\mu_2 \cos\psi]$$

$$+a_{21}[\sin\phi \cos\mu_1 - \sin\mu_1 \cos\phi][cos\psi \cos\mu_2 + \sin\psi \sin\mu_2]$$

$$+a_{22}[\sin\phi\cos\mu_1 - \sin\mu_1\cos\phi][\sin\psi\cos\mu_2 - \sin\mu_2\cos\psi]$$

$$= \kappa_1\cos(\phi) + \kappa_2\cos(\psi) + a_{11}\cos(\phi)\cos(\psi) + a_{22}\sin(\phi)\sin(\psi).$$

Now, we approximate this quantity at second order in the limit of $(\phi, \psi) \simeq (0, 0)$ and get

$$\log(f_{VM}(\phi,\psi)) \propto \kappa_1\left(1 - \frac{\phi^2}{2}\right) + \kappa_2\left(1 - \frac{\psi^2}{2}\right)$$

$$+a_{11}\left(1 - \frac{\phi^2}{2}\right)\left(1 - \frac{\psi^2}{2}\right) + a_{22}\left(1 - \frac{\phi^2}{2}\right)\left(1 - \frac{\psi^2}{2}\right).$$

Since we are approximating at second order in (ϕ, ψ), we get

$$\log(f_{VM}(\phi,\psi)) \propto -\kappa_1\frac{\phi^2}{2} - \kappa_2\frac{\psi^2}{2} - a_{11}\frac{\phi^2}{2} - a_{11}\frac{\psi^2}{2} - a_{22}\frac{\phi^2}{2} - a_{22}\frac{\psi^2}{2}.$$

We do the same for the *Inverse Stereographic Bivariate Normal Distribution* and obtain

$$-2log(f_{ISN})$$

$$\propto b_{11}\left(\frac{\sin\phi}{1+\cos\phi} - \mu\right)^2 + 2b\left(\frac{\sin\phi}{1+\cos\phi} - \mu\right)\left(\frac{\sin\psi}{1+\cos\psi} - \nu\right) + b_{22}\left(\frac{\sin\psi}{1+\cos\psi} - \nu\right)^2$$

$$\simeq b_{11}\left(\frac{\phi}{2} - \mu\right)^2 + 2b\left(\frac{\phi}{2} - \mu\right)\left(\frac{\psi}{2} - \nu\right) + b_{22}\left(\frac{\psi}{2} - \nu\right)^2$$

$$= b_{11}\left(\frac{\phi^2}{4} - \phi\mu + \mu^2\right) + 2b\left(\frac{\phi\psi}{4} - \nu\frac{\phi}{2} - \nu\frac{\psi}{2} + \mu\nu\right) + b_{22}\left(\frac{\psi^2}{4} - \psi\nu + \nu^2\right)$$

$$= -b_{11}\frac{\phi^2}{2} - b_{22}\frac{\psi^2}{2}.$$

Here, $b_{ij} := (\Sigma^{-1})_{ij}$ for $i, j = 1, 2$. The theorem then follows by matching the coefficients and by the fact that our Taylor expansion stopped at second order in (ϕ, ψ). □

Remark 4.5.24. *Note that some choices are not possible; for example, the choice* $\mu_1 = \mu_2 = 0$ *gives some constraints on the parameters* $\mu, \nu, b_{11}, b, b_{22}$*.*

4.5.4 Inverse Stereographic Moments, Inverse Stereographic Moment Generating Function and a version of the Central Limit Theorem

In this section, we introduce the *Inverse Stereographic Moments* and the *Inverse Stereographic Moment Generating Function.* We use them to rephrase the Central Limit Theorem in the context of $\mathbb{T}^n$.

4.5.4.1 The Inverse Stereographic Moments

In this subsubsection, we introduce a suitable notion of *Moments* and *Moment Generating Function* which is in agreement with the way in which we have constructed the densities on the *Circle* and *Torus*, namely, through the *Stereographic Projection.*

We start by describing how we modify the classical *Circular* and *Spherical Moments* in order to make them natural for the *Stereographic Projection.* At this point, our discussion sticks to $\mathbb{S}^1$ to simplify the explanation, but later we will define the *Inverse Stereographic Moments* and *Inverse Stereographic Generating Function* for the *Spherical* and the *Toroidal* cases as well.

Definition 4.5.25. *Consider a random variable Θ defined on $\mathbb{S}^1$ with pdf given by $f_\Theta(\theta)$, for every $\theta \in [-\pi, +\pi)$. Then, the* Circular Moments of Θ *are defined as*

$$m_n := E(z^n) := \int_{[-\pi,+\pi)} f_\Theta(\theta) e^{in\theta} d\theta,$$

where $z = e^{i\theta}$.

Moreover, we can define the following quantities: the *Population Resultant Vector* $\rho := m_1$; the *Length* $R := |m_1|$; the *Mean Angle* $\theta_\mu := \mathrm{Arg}(m_1)$; the *Lengths of the Higher Moments* $R_n := |m_n|$; the *Angular Parts of the Higher Moments* $(n\theta_\mu) \mod 2\pi$.

Remark 4.5.26. *Note that, as an easy consequence of Hölder's Inequality with $p = 1$ and $q = \infty$,* Circular Moments *are well defined for any random variable Θ which takes values on $\mathbb{S}^1$.*

We can define *Sample Moments* analogously.

Definition 4.5.27. *Suppose we have a set of observations $\theta_1, \ldots \theta_n$ on the random variable*

Θ. *Then, we can define the* Sample Moments *of order n as*

$$\overline{m}_n := \frac{1}{N}\sum_{i=1}^{N} z_i^n.$$

Moreover, we can define the following quantities: the *Sample Resultant Vector* $\overline{\rho} := \frac{1}{N}\sum_{n=1}^{N} z_n$; the *Sample Mean Angle* $\overline{\theta} := \mathrm{Arg}(\overline{\rho})$; the *Length of the Sample Mean Resultant Vector* $\overline{R} := |\overline{\rho}|$.

Remark 4.5.28. *A simple consequence of the previous definitions is that the* Sample Mean Resultant Vector *can be represented as*

$$\overline{\rho} = \overline{R}\, e^{i\overline{\theta}}.$$

This is the classical way in which *Moments on the Circle* are defined. Roughly speaking, one considers the *Lebesgue Measure* $d\theta$ on $[-\pi, +\pi)$, restricts polynomials z^n to the *Circle* $\mathbb{S}^1$ and computes the *Moments* accordingly. This operation is legitimate and proved to be useful to answer statistical questions.

However, this procedure has no connection with the *Stereographic Projection*, since the *Stereographic Projection* does not send the *Lebesgue Measure on* $[-\pi, +\pi)$ (and the corresponding the Circle) to the *Lebesgue Measure on the Real Line* and also does not send polynomials defined on $\mathbb{C} \simeq \mathbb{R}^2$ to their restriction to $\mathbb{S}^1$.

So, we propose to compute *Moments* in a way which is consistent with the *Stereographic Projection*. Therefore, this way of computing Moments is particularly suitable for random variables defined on the circle obtained by *Inverse Stereographic Projection*. This procedure identifies in a geometrically natural way the circular counterpart of the moments defined in the *Euclidean Space* with respect to the *Lebesgue Measure*. A corresponding construction would also work in higher dimensions.

Definition 4.5.29. *Consider a random variable Θ defined on $\mathbb{S}^1$ with pdf $f_\Theta(\theta)$. Then, we define the k-*Inverse Stereographic Circular Moments *as*

$$m_k^{\mathbb{S}^1}[\Theta] := E_S\left[\left(\frac{\sin\theta}{1+\cos\theta}\right)^k\right] := \int_{-\pi}^{+\pi}\left(\frac{\sin\theta}{1+\cos\theta}\right)^k f_\Theta(\theta)d\theta.$$

Remark 4.5.30. *Note that, by applying the* Stereographic Projection*, we get*

$$m_k^{\mathbb{S}^1}[\Theta] = \int_{-\infty}^{+\infty} x^k f_\Theta(2\arctan(x))\frac{2dx}{1+x^2}.$$

Therefore, if $f_\Theta(\theta)$ *comes from an* Inverse Stereographic Projection

$$m_k^{\mathbb{S}^1}[\Theta] = \int_{-\infty}^{+\infty} x^k f_X(x)dx$$

with X *being the image of* Θ *under* Stereographic Projection*.*

Remark 4.5.31. *With this definition of* Moments of a Random Variable*, not all random variables defined on the* Circle *admit finite* Inverse Stereographic Moments *(Hölder's Inequality is not conclusive in this case). For example, the* Uniform Distribution *on* $\mathbb{S}^1$ *(*$\Theta \sim Unif(0, 2\pi)$*), which corresponds to the* Cauchy Distribution on the Real Line *through* Inverse Stereographic Projection*, does not admit any finite* Inverse Stereographic Moment*.*

This definition can be easily extended to random variables defined on $\mathbb{S}^n$ and $\mathbb{T}^n$. We will just discuss the case of the torus, since it is of interest in the subsequent discussion.

Definition 4.5.32. *Consider a random variable* $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_n)$ *defined on* $\mathbb{T}^n := \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ *with pdf given by* $f_{\mathbf{\Theta}}(\theta)$ *and* $\theta := (\theta_1, \ldots, \theta_n)$*. Then, we define the* $(k_1, \ldots, k_n)$*-* Inverse Stereographic Toroidal Moment *as*

$$m_{(k_1,\ldots,k_n)}^{\mathbb{T}^n}[\mathbf{\Theta}] := E_{\mathbb{T}^n}\left[\left(\frac{\sin\theta_1}{1+\sin\theta_1}\right)^{k_1}, \ldots, \left(\frac{\sin\theta_n}{1+\sin\theta_n}\right)^{k_n}\right]$$

$$:= \int_{[-\pi,+\pi)^n} \prod_{i=1}^n d\theta_i \left[\left(\frac{\sin\theta_1}{1+\sin\theta_1}\right)^{k_1} \times \cdots \times \left(\frac{\sin\theta_n}{1+\sin\theta_n}\right)^{k_n}\right] f_{\mathbf{\Theta}}(\theta).$$

Here, $k_1, \ldots, k_n \in \mathbb{N}$*.*

Example 4.5.33. *With this definition, it is easy to see that the* Inverse Stereographic Moments *of a random variable distributed as* $\mathcal{MISN}(\mu, \mathbf{\Sigma})$ *coincide with the standard moments of the* Multivariate Normal Distribution *in* $\mathbb{R}^n$ *and so*

$$E^{\mathbb{T}^n}[\mathcal{MISN}(\mu, \mathbf{\Sigma})] = \mu$$

and

$$Var^{\mathbb{T}^n}[\mathcal{MISN}(\mu, \mathbf{\Sigma})] = \mathbf{\Sigma}.$$

Remark 4.5.34. *An important consequence of this definition of moments is that we can compute moments of distributions coming from* Inverse Stereographic Projections *analytically and by performing simple integrations on* $\mathbb{R}^n$*. This is a big simplification also from the point of view of the implementation, since the estimation of parameters would not need numerical optimization algorithms, in general. Note that the simplification in the computation is directly related to how simple the computation of the moments is in the corresponding random variable on the Euclidean space.*

Remark 4.5.35. *These definitions of moments turn out to be somehow independent of the choice of the* North Pole*. Think about the case* $n = 2$*. The choice of the* North Pole N *is as arbitrary as the choice of the origin or the point at infinity in the real line. A line does not know anything about the system of coordinates that we put on. In the same way, the circle has no intrinsically well defined north pole. However, we want that characteristic quantities like the expected value and the variance are as independent as possible with respect to the choice of the* North Pole*.*

If you choose a north pole $N' \neq N$*, you can go from one* Stereographic Projection *to another by a simple change in the angles* $\theta_i \mapsto \theta + \theta_i^0$ *for* $i = 1, \ldots, n$*. This choice of* N' *gives us a new* Inverse Stereographic Projection *and so a new pdf* $f_\Theta(\theta + \theta_0) \neq f_\Theta(\theta)$*. The point is that, the measure also changes consistently and so also* $\frac{d\theta}{1+\cos\theta} \mapsto \frac{d\theta}{1+\cos(\theta+\theta_0)}$*, which compensates the change in the pdf.*

This procedure of choosing a geometric definition of moments leaves every characteristic quantity to be invariant. There is still the dependence on the choice of the origin in the Euclidean Space and so the definition is not perfectly independent of any coordinate system. However, this sort of dependence on the coordinate system is the same as the Euclidean one, where it is well accepted.

4.5.5 A Central Limit Theorem and the Inverse Stereographic Moment Generating Function

In this subsection, we construct an *Inverse Stereographic Moment Generating Function* that we use to prove a version of the *Central Limit Theorem on the Circle*.

We do not prove the most general version possible and we will stick to the one dimensional case, even though the theorem works in higher dimensions as well. The result is a consequence of the definition of *Inverse Stereographic Moment Generating Function* and of the *Central Limit Theorem on the Euclidean Space.*

Definition 4.5.36. *Consider a random variable* Θ *defined on* $\mathbb{S}^1$ *with pdf* $f_\Theta(\theta)$*. Then, we define the* Inverse Stereographic Circular Moment Generating Function *as*

$$M^{\mathbb{S}^1}[\Theta] := E_S\left[e^{t\tan\frac{\theta}{2}}\right] := \int_{-\pi}^{+\pi} e^{t\tan\frac{\theta}{2}} f_\Theta(\theta)d\theta.$$

Remark 4.5.37. *Note that, by applying the* Stereographic Projection*, we get*

$$M_n^{\mathbb{S}^1}[\Theta] = \int_{-\infty}^{+\infty} e^{tx} f_\Theta(2\arctan(x))\frac{2dx}{1+x^2}.$$

Therefore, if $f_\Theta(\theta)$ *comes from an* Inverse Stereographic Projection*,*

$$M^{\mathbb{S}^1}[\Theta] = \int_{-\infty}^{+\infty} e^{tx} f_X(x)dx$$

with X *being the image of* Θ *under* Stereographic Projection*.*

This makes the above choice of moments a natural one. We can construct a *Inverse Stereographic Circular Moment Generating Function* in higher dimensions as well.

We concentrate on the case of the *Torus.*

Definition 4.5.38. *Consider a random variable* $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_n)$ *defined on* $\mathbb{T}^n := \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ *with pdf given by* $f_{\mathbf{\Theta}}(\theta)$ *and* $\theta := (\theta_1, \ldots, \theta_n)$*. Then, we define the* Inverse Stereographic Toroidal Moment Generating Function *as*

$$M_{\mathbf{\Theta}}^{\mathbb{T}^n}(t_1, \ldots, t_n) := E_{\mathbb{T}^n}\left[e^{\sum_{i=1}^n t_i \tan\frac{\theta_i}{2}}\right] := \int_{[-\pi,+\pi)^n} \prod_{i=1}^n d\theta_i e^{\sum_{i=1}^n t_i\tan\frac{\theta_i}{2}} f_{\mathbf{\Theta}}(\theta).$$

Here, $t_1, \ldots, t_n \in \mathbb{R}$*.*

We are now ready to state and prove a version of the *Central Limit Theorem* suitable for *Inverse Stereographic Probability Distributions.*

Theorem 4.5.39. *Suppose* $\Theta_1, \Theta_2, \ldots$ *is a sequence of iid random variables which take values in* $\mathbb{S}^1$ *and which are inverse stereographic projected by some random variables* $X_1, X_2, \ldots$ *on the real line. Suppose the* Inverse Stereographic Circular Moment Generating Function *exists in a neighbourhood of* $t = 0$*. Further, suppose that* $E^{\mathbb{S}^1}[\Theta_i] = \mu$

and $Var[\Theta_i] = \sigma^2 > 0$. *Define*

$$\Phi_n := P^{-1}\left(\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} P(\Theta_i) - \mu\right)}{\sigma}\right),$$

with P *being the* Stereographic Projection. *Then,*

$$\Phi_n \to \mathcal{ISN}(0,1) \text{ as } n \to +\infty,$$

in the sense of the Inverse Stereographic Circular Moment Generating Function *and so in distribution.*

Remark 4.5.40. *This is not the most general version of* Central Limit Theorem *that we can prove. In particular, the condition on the existence of the stereographic mgf is not necessary, as in the usual case. However, we stick to this formulation for the simplicity of the argument.*

Proof. First of all, notice that

$$M_{\Theta_i}^{\mathbb{S}^1}(t) = M_{P(\Theta_i)}(t),$$

for every $i = 1, \dots, n$. This implies that

$$E^{\mathbb{S}^1}[\Theta_i] = \mu, \qquad Var^{\mathbb{S}^1}[\Theta_i] = \sigma^2.$$

Now, by definition of *Inverse Stereographic Moment Generating Function*, we obtain

$$M_{\Phi_n}^{\mathbb{S}^1}(t) = M^{\mathbb{S}^1}_{P^{-1}\left(\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} P(\Theta_i)-\mu\right)}{\sigma}\right)}(t) = M_{\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} P(\Theta_i)-\mu\right)}{\sigma}}.$$

The random variable $\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} P(\Theta_i)-\mu\right)}{\sigma}$ satisfies all the hypotheses of the classical *Central Limit Theorem* and, in particular, independence is preserved by the *Stereographic Projection* and this can be seen by the following chain of equalities:

$$f_{\Theta_1,\Theta_2}(\theta_1,\theta_2) = f_{X_1,X_2}\left(\frac{\sin\theta_1}{1+\cos\theta_1}, \frac{\sin\theta_2}{1+\cos\theta_2}\right)\frac{d\theta_1}{1+\cos\theta_1}\frac{d\theta_2}{1+\cos\theta_2}$$

$$= f_{X_1}\left(\frac{\sin\theta_1}{1+\cos\theta_1}\right) f_{X_2}\left(\frac{\sin\theta_2}{1+\cos\theta_2}\right)\frac{d\theta_1}{1+\cos\theta_1}\frac{d\theta_2}{1+\cos\theta_2} = f_{\Theta_1}(\theta_1)f_{\Theta_2}(\theta_2).$$

Therefore,

$$\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}P(\Theta_i)-\mu\right)}{\sigma}\to\mathcal{N}(0,1)$$

in distribution. This implies that

$$M_{\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}P(\Theta_i)-\mu\right)}{\sigma}}\to M_{\mathcal{N}(0,1)}=M_{P^{-1}(\mathcal{N}(0,1))}(t)=M_{M_{\mathcal{ISN}(0,1)}}(t).$$

Therefore,

$$M_{\Phi_n}^{\mathbb{S}^1}(t)\to M_{M_{\mathcal{ISN}(0,1)}}(t)\ \textit{as}\ n\to+\infty,$$

and so

$$\Phi_n\to\mathcal{ISN}(0,1)\ \textit{as}\ n\to+\infty.$$

This conclues the proof of the theorem. □

Remark 4.5.41. *This version of the* Central Limit Theorem *is a straightforward consequence of the classical* Central Limit Theorem *and the smoothness of the* Stereographic Projection*.*

4.5.6 Inference

In this subsection, we discuss some key inferential issues such as point estimation, confidence intervals and hypothesis testing. We also discuss some sampling methods for inverse stereographic probability distributions.

4.5.6.1 Parameter Estimation and Hypothesis Testing

The first statistical problem that we address is the problem of *Parameter Estimation*. We need to find good *Statistics* or *Estimators* which take values in $\mathbb{S}^1$, $\mathbb{S}^n$, $\mathbb{T}^n$ or any other manifold which can be flattened out on the *Euclidean Space* through *Stereographic Projection*.

Definition 4.5.42. *Consider a* statistic T_X *on* $\mathbb{R}^p$*. Then, we call* Inverse Stereographic Statistic*, the following real-valued or vector-valued function:* $T_\Theta := P^{-1}\circ T_X\circ P$*, where* P *is the* Stereographic Projection*.*

Remark 4.5.43. *These estimators are natural for distributions defined through* Inverse Stereographic Projection *and, in general, this definition works well, whenever we transport a random variable from one space to another through an invertible map* P*.*

Example 4.5.44. *Suppose we have a random sample* $\Theta_i\sim\mathcal{ISN}(\mu,\sigma^2)$*, for* $i=1,\ldots,n$*. Consider the* Stereographic Projection P*. We want to define the* Inverse Stereographic

Sample Mean *and then compute its distribution. The* Inverse Stereographic Sample Mean *can be defined as*

$$\bar{\Theta}^S := P^{-1}\left(\sum_{i=1}^{n} \frac{P(\Theta_i)}{n}\right),$$

where $\Theta_i \sim \mathcal{ISN}(\mu, \sigma^2)$*, for* $i = 1, \ldots, n$*. By the definition of* P *and since* $\Theta_i \sim \mathcal{ISN}(\mu, \sigma^2)$*, for* $i = 1, \ldots, n$*, we have* $P(\Theta_i) \sim N(\mu, \sigma^2)$*, for* $i = 1, \ldots, n$*, and so* $\sum_{i=1}^{n} \frac{P(\Theta_i)}{n} \sim N(\mu, \sigma^2/n)$*. Again, by the definition of* P*, we have* $\bar{\Theta}^S \sim \mathcal{ISN}(\mu, \sigma^2/n)$*.*

Note that similar considerations can be done for the Sample Variance *and other estimators, and that* $P(\bar{\Theta}^S)$ *is a point estimator of* μ*.*

Remark 4.5.45. *The canonical definition of circular moments run into some problems due to the fact that the parameters that one uses in the circular cases mimic the ones used in the Euclidean case and do not take into consideration the different geometry of the circle, sphere and torus, for example.*

Consider the sample mean of two angles. If the angles are $\theta_1 = -1$ *and* $\theta_2 = +1$*, one has no problem to say that the mean is* 0*. Suppose, instead, the angles are* $-\frac{\pi}{2}$ *and* $\frac{\pi}{2}$*. Then, it is not clear if it is more reasonable to say that the mean is* 0 *or* π*.*

The knowledge of the mean is much more meaningful in the Euclidean setting than in a periodic setting, where "$-\infty$ *and* $+\infty$ *join together". The* Inverse Stereographic Moments *address this issue by breaking the symmetry of the circle and keeping* $-\infty$ *and* $+\infty$ *distinct. In this way, the problem mentioned above (of a not unequivocal definition of mean) no longer exists. This procedure also suggests that it might be more intuitive to give an interpretation of parameters of circular/spherical/toroidal distributions after stereographic projection of the distribution.*

We can also develop *Interval Estimation* for random variables distributed as an *Inverse Projected Distributions*. Now, we concentrate on *Confidence Intervals*. *Confidence Intervals* are *Interval Estimators* with a measure of confidence. We call $(1-\alpha)$-*Confidence Interval* a *Confidence Interval* with a *Confidence Coefficient* equal to $1-\alpha$. We construct a *Confidence Interval* for the *Population Mean* of a random variable distributed as a *Inverse Stereographic Normal*.

Example 4.5.46. *Suppose we have a random variable* $\Theta \sim \mathcal{ISN}(\mu, \sigma^2)$ *defined on* $\mathbb{S}^1$ *with* μ *unknown and* σ^2 *known. Consider the* Stereographic Projection P*. We want to* *construct a* $(1-\alpha)$-Confidence Interval for μ. *A* Confidence Interval *for* μ *depends on* statistics *of the form*

$$[L(\theta_1,\ldots,\theta_n), U(\theta_1,\ldots,\theta_n)].$$

It is well known that a confidence interval for μ *in the case of* Normal distribution *is of the form:*

$$C(x_1,\ldots,x_n) := \left\{\mu : \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\},$$

where n *is the sample size,* $x_1,\ldots,x_n$ *is the sample realization,* $\bar{x}$ *is the sample mean and* $z_{\alpha/2}$ *is the* quantile *of the* Standard Normal Distribution*.*

Now, we construct a corresponding confidence interval on $\mathbb{S}^1$ *at level* $1-\alpha$*. We have*

$$1-\alpha = Pr\left(c \leq P\left(\bar{\Theta}^S\right) - \mu \leq d\right)$$

$$= Pr\left(c+\mu \leq P\left(\bar{\Theta}^S\right) \leq d+\mu\right) = Pr\left(c+\mu \leq \frac{1}{n}\sum_{i=1}^{n} P(\Theta_i) \leq d+\mu\right)$$

$$= Pr\left(\frac{c}{\sigma/\sqrt{n}} \leq Z \leq \frac{d}{\sigma/\sqrt{n}}\right) = \Phi(d\sqrt{n}/\sigma) - \Phi(c\sqrt{n}/\sigma),$$

with $Z \sim N(0,1)$*. A possible choice is then* $c = -d = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$*. Therefore,*

$$C(\theta_1,\ldots,\theta_n) := \left\{\mu : P(\bar{\theta}^S) + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq P(\bar{\theta}^S) - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

becomes an Inverse Stereographic Interval Estimator of μ. *Since*

$$P(C(\theta_1,\ldots,\theta_n)) = 1-\alpha$$

for some $\alpha \in [0,1]$*, then* $C(\theta_1,\ldots,\theta_n)$ *is a* $(1-\alpha)$-Inverse Stereographic Confidence Interval of μ*.*

Remark 4.5.47. *Note that* $\bar{\theta}^S \sim \mathcal{ISN}(0,\sigma^2)$ *if and only if* $P(\bar{\theta}^S) \sim N(0,\sigma^2)$ *and so the confidence interval does not depend on* P*.*

Remark 4.5.48. *Note that there are two fundamental advantages in using this perspective: a theoretical and a practical one. From a theoretical point of view, everything is geometrically consistent. From a practical point of view, the estimators are explicit and their distributions can be computed as explicitly as they can be computed on the* Euclidean Space*. Therefore, there is no need for numerical optimization, when there is no need in the Euclidean counterparts.*

Now, we explain the construction of *Hypothesis Tests*.

Example 4.5.49. *Suppose we have a random variable $\Theta \sim \mathcal{ISN}(\mu, \sigma^2)$ defined on $\mathbb{S}^1$ with μ unknown and σ^2 known. Consider the* Stereographic Projection P. *Suppose we want to test the following hypotheses:*

$$H_0 : \mu = \mu_0 \text{ vs } H_a : \mu \neq \mu_0.$$

Consider a random sample $\Theta_1, \ldots, \Theta_n \sim \mathcal{ISN}(\mu, \sigma^2)$. We can use the inversion theorem and build a Rejection Region from the Confidence Intervals described above. For a fixed level α, the Most Powerful Unbiased Test *(see [25]) has rejection region*

$$R(\theta) := \left\{\theta = (\theta_1, \ldots, \theta_n) : |P(\bar{\theta}^S) - \mu| > z_{\alpha/2}\sigma/\sqrt{n}\right\}.$$

This test has size α if $P(H_0$ is rejected $|\mu = \mu_0) = \alpha$.

4.5.7 Sampling Methods for Inverse Stereographic Projected Distributions

The aim of the next few subsections is to show how easy it is to sample from *Inverse Stereographic Projected Distributions*. The method is to sample from the corresponding Euclidean Distributions and then inverse stereographic project the sample on the circle, sphere or torus.

Due to the applications in molecular biology that we discuss later, we concentrate on the case of the torus $\mathbb{T}^2 := \mathbb{S}^1 \times \mathbb{S}^1$ and on the *Inverse Stereographic Normal Distribution*.

The method we use to sample from a random variable distributed as ISND is the so called *Box-Muller Transformation* (see [111]).

<u>The Box-Muller Transformation</u>

The *Box-Muller Sampling* is based on representing in polar coordinates the joint distribution of two independent *Standard Normal Random Variables* X and Y:

$$X \sim N(0, 1) \qquad Y \sim N(0, 1).$$

The joint distribution $f_{X,Y}(x, y)$ is

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}},$$

and is invariant by rotations of the (x, y)-plane. Note that $x^2+y^2 = r^2$ and this suggests to represent/transform the pdf of the *Bivariate Normal Distribution* in/to polar coordinates as

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{\frac{-r^2}{2}}.$$

This implies that the joint pdf of r and θ is given by

$$f_{r,\theta}(r, \theta) = \frac{1}{2\pi} e^{\frac{-r^2}{2}} r.$$

From this, it is clear that $f_{r,\theta}(r, \theta)$ is the product of two density functions: *Exponential Distribution* of the square radius

$$r^2 \sim Exp\left(\frac{1}{2}\right);$$

and the *Uniform Distribution* of the angle

$$\theta \sim Unif(0, 2\pi).$$

Recall the connection between the exponential distribution and the uniform distribution, given by

$$Exp(\lambda) = \frac{-\log(Unif(0, 1))}{\lambda}.$$

Then,

$$r \sim \sqrt{-2\log(Unif(0, 1))}$$

and this gives us a way to generate points from the *Bivariate Normal Distribution* by sampling from two independent *Uniform Distributions*, one for the radius r and another for the angle θ.

The algorithm goes as follows:

- Draw $U_1, U_2 \sim Unif(0, 1)$;
- Transform the variables into radius and angle by $r = \sqrt{-2\log(U_1)}$, and $\theta = 2\pi U_2$;
- Transform radius and angle back to

$$x = r\cos\theta, \quad y = r\sin\theta.$$

The result of this procedure is the generation of two independent Normal Random Variables X and Y, based on the ability to generate $U_1, U_2 \sim Unif(0,1)$.

Remark 4.5.50. *We have assumed here, of course, that we can generate iid* Uniform Random Variables*. This is not a trivial problem, but this has been discussed extensively in the literature. For a nice review of several existing methods of generating the* Uniform Distribution *on various* Compact Spaces*, we refer to [39], [40] and [92].*

A simple MATLAB implementation of the *Box-Muller algorithm* is shown below:

```
% NORMAL SAMPLES USING BOX-MUELLER METHOD
u = rand(2,100000);
r = sqrt(-2*log(u(1,:)));
theta = 2*pi*u(2,:);
x = r.*cos(theta);
y = r.*sin(theta);

% DISPLAY BOX-MULLER SAMPLES
figure
% X SAMPLES
subplot(121);
hist(x,100);
colormap hot;axis square
title(sprintf('Box-Muller Samples Y\n Mean = %1.2f\n Variance = %1.2f\n
 Kurtosis = %1.2f',mean(x),var(x),3-kurtosis(x)))
xlim([-6 6])

% Y SAMPLES
subplot(122);
hist(y,100);
colormap hot;axis square
title(sprintf('Box-Muller Samples X\n Mean = %1.2f\n Variance = %1.2f\n
 Kurtosis = %1.2f',mean(y),var(y),3-kurtosis(y)))
xlim([-6 6])
```

At this point, one needs only to inverse stereographic project the random variables X and Y just obtained to get the corresponding ISND on $\mathbb{S}^1$.

4.5.8 Applications and Numerical Examples

In this subsection, we give some numerical examples and applications of the theory developed in the preceding sections.

4.5.8.1 Plots and Animations

In this subsubsection, we collect some plots and animations in order to visualize the *Inverse Stereographic Normal Distribution* and to see how the parameters change the location, scale and shape of the distribution.

We can see that the distribution can be unimodal and multimodal, depending on the range of parameters that we have used. Also, we underline that the pdf in Figure 1 is unimodal and the "multimodal looking" behaviour is just apparent. In fact, the sides of the square of the *Ramachandran Plot* (see [97]) are identified and so the four angles are actually the same point. The graph is in agreement with Theorem 4.5.17 and Theorem 4.5.19. Note how simpler it is to simulate ISDN with respect to simulate Von Mises distributions [14].

```
scale = [linspace(0,10,200)];    % surface scaling (0 to 10, 200 frames)
for ii = 1:length(scale)
x1 = linspace(-pi, pi);
x2 = linspace(-pi, pi);
[X1,X2]= meshgrid(x1,x2);
Y1 = sin (X1)./(1+cos(X1));
Y2 = sin (X2)./(1+cos(X2));
c=scale(ii);
Z1=(2*pi)^(-0.5)*exp(-0.5.*c*(Y1).^2)./(c*(1+cos(X1)));
Z2=(2*pi)^(-0.5)*exp(-0.5.*c*(Y2).^2)./(c*(1+cos(X2)));
Z=Z1.*Z2;
surf(X1,X2,Z)
pause(0.05)     %  control of the animation speed
end


% Stereographic NORMAL SAMPLES USING BOX-MUELLER METHOD
% DRAW SAMPLES FROM PROPOSAL DISTRIBUTION
u = rand(2,1000);
r = sqrt(-2*0.1*log(u(1,:)));
theta = 2*pi*u(2,:);
```

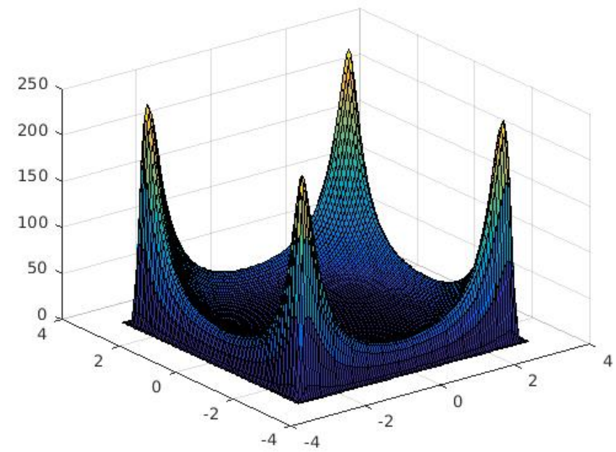


Figure 4.1: Parameter $\sigma = 0.1$

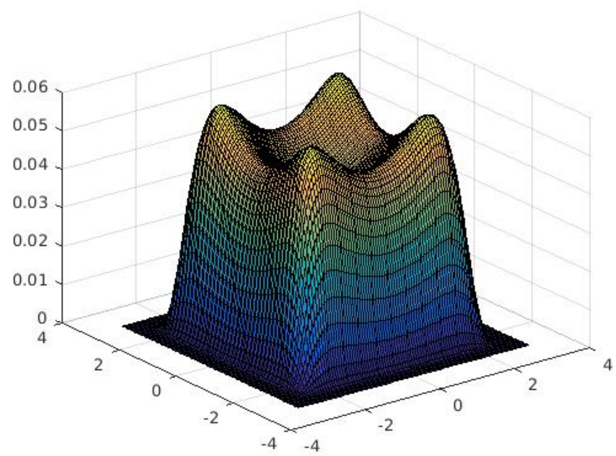


Figure 4.2: Parameter $\sigma = 1$

```
x = r.*cos(theta);
y = r.*sin(theta);
theta1=2*atan(x-1);
theta2=2*atan(y-1);

figure
xlim([-pi pi])
ylim([-pi pi])
```

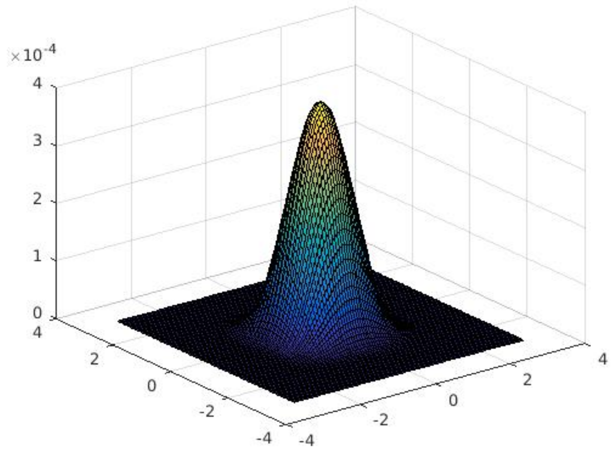


Figure 4.3: Parameter $\sigma = 10$

```
scatter(theta1,theta2)
```

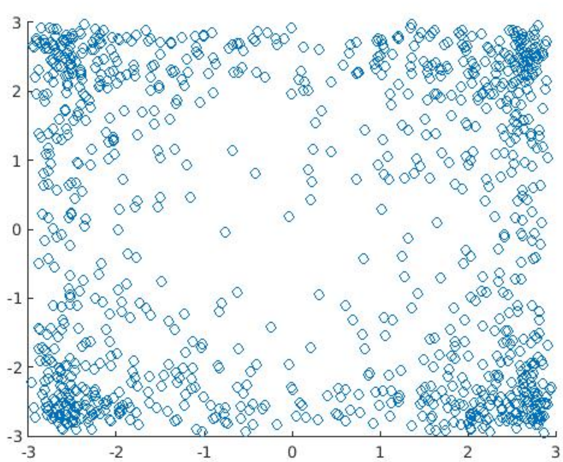


Figure 4.4: Parameter $\sigma = 0.1$, $\mu = 0$

4.5.8.2 Comparison between the von Mises distribution and the ISND

In this subsubsection, we use Theorem 4.5.21 to show how in the case of high-concentration limit the BVM distribution is well approximated by the *Bivariate Inverse Stereographic Normal Distribution* (BISND) for a certain range of parameters.

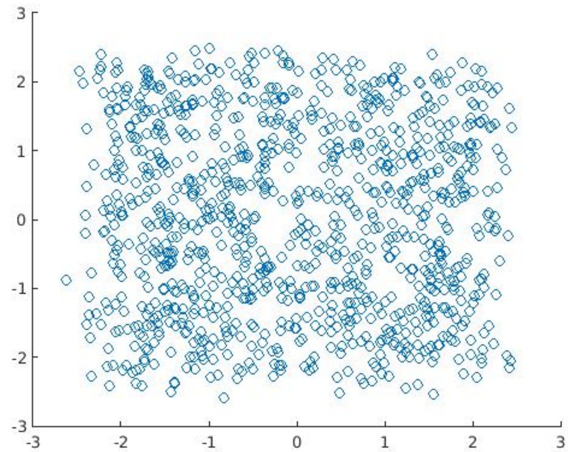


Figure 4.5: Parameter $\sigma = 1$, $\mu = 0$

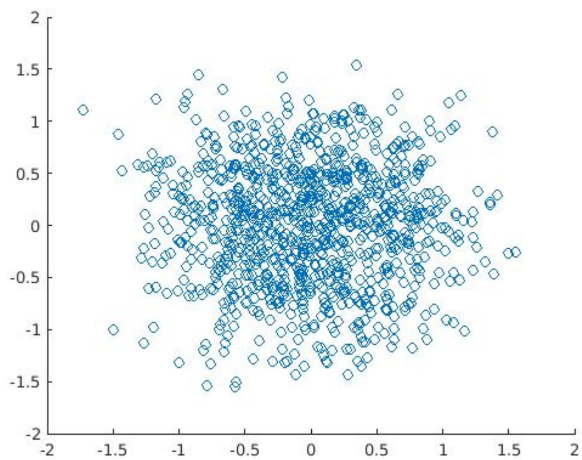


Figure 4.6: Parameter $\sigma = 10$, $\mu = 0$

Consider a BVM distribution with parameters $\nu = \mu = 0$, $\mathbf{A} = Id_{2\times 2}$ and $\kappa_1 = \kappa_2 = 100$. Then, by using Theorem 4.5.21, we can find some parameters of the BISND that approximate the BVM up to third order. By Theorem 4.5.21, we get: $\mu_1 = \mu_2 = 0$, $b_{11} = b_{22} = -101$, $b = -1$. We can see by the following R-plot (Figure 8) how close the two curves are: the black one is the BISND, while the red one is the BVM.

```
#Comparison von Mises ISND
```

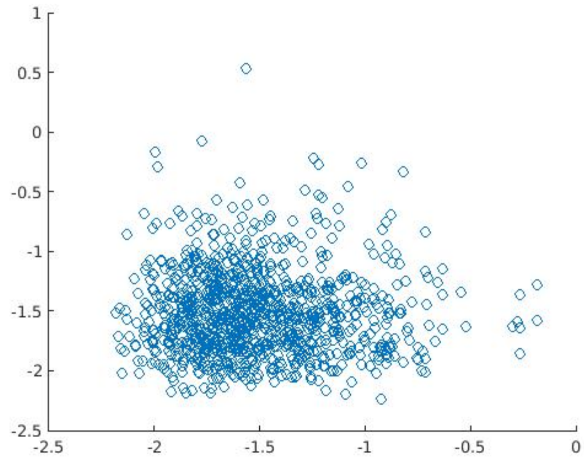


Figure 4.7: Parameter $\sigma = 10$, $\mu = 1$

```
curve(sqrt(101)/(sqrt(2*pi)*(1+cos(x+pi/2)))
*exp(-0.5*101*(sin(x+pi)/(1+cos(x+pi)))^2),
 from=-pi, to=pi, xlab="", ylab="")
par(new = TRUE)
curve((1/(2*pi*besselI(100, 0, expon.scaled = FALSE)))
*exp(100*cos(x-pi)),
 from=-pi, to=pi, xlab="", ylab="", axes=FALSE, col="red")
```

We can fit a BISND to the data analysed in [110] in a similar way. In that section, the authors imposed $a_{11} = a_{12} = a_{21} = 0$ and found $\hat{\kappa}_1 = 35.41$, $\hat{\kappa}_1 = 20.17$, $\mu = 0.073$ rad, $\nu = -1.560$ rad and $\hat{a}_{22} = -13.70$. For the corresponding parameters in the BISND, we get $\mu_1 = 0.05903399$, $\mu_2 = 2.123424$, $b_1 1 = 34.66034$, $b = -7.178959$, $b_{22} = -16.525$. These numbers seem completely different from the ones obtained for the von Mises fit, and this is not surprising, since the hypothesis of Theorem 4.5.21 are not satisfied in this case.

4.5.8.3 Goodness of Fit Tests

In this subsubsection, we discuss the problem of *Goodness of Fit*. We want assess whether the model that we have proposed is indeed a good model for the observed data. Later, we will test if the distribution of points follows a ISND or another distribution, using these type of tests.

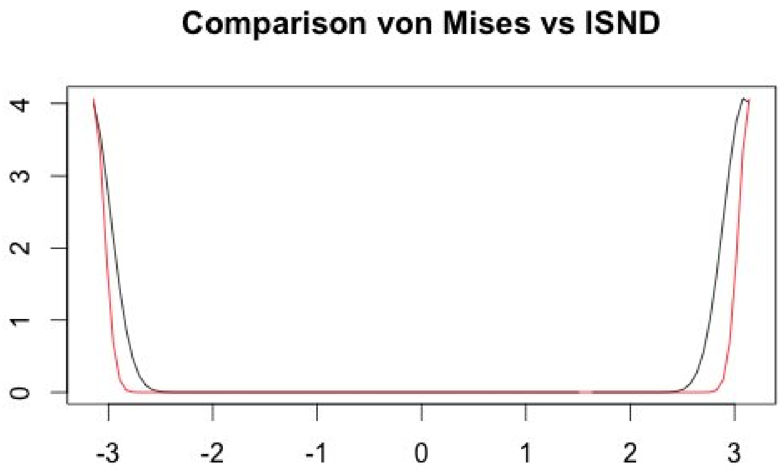


Figure 4.8: The black curve is the Bivariate Inverse Stereographic Normal. The red curve is the Bivariate von Mises

Again, due to the duality given by the *Stereographic Projection*, we reduce the problem to the *Goodness of Fit Tests* on the *Euclidean Space*. Of course, there are several methods to test normality of a dataset or any other type of distributional form, but since we want to provide an illustration of the potential of the method, we just stick to some of the most common ones, like the *Chi-Squared Goodness of Fit Test*, the *Shapiro-Wilk Test*, and the *Kolmogorov-Smirnov Test*.

Now, we show a couple of examples as to how the tests work in a case in which we expect rejection and also in a case in which we expect to not reject.

Example 4.5.51. *If we sample uniform distributed points on the circle, we can see that a simple* Chi-square Test *and a simple* Shapiro-Wilk Test *([107], [100], [101], [102] and [103]) reject the null hypothesis H_0 of being* Inverse Stereographic Normal*.*

```
> #Uniformity Test
> set.seed(100)
> theta<-c(runif(100, min = -pi, max = pi));
> x<-tan(theta/2);
> shapiro.test(x)
```

```
Shapiro-Wilk normality test

data:  x
W = 0.5798, p-value = 1.811e-15
```

Example 4.5.52. *If we instead sample from a* $\mathcal{ISND}(\mu, \Sigma)$*, it is enough to test for normality on the Euclidean Space:*

```
> set.seed(100)
> shapiro.test(rnorm(5000, mean = 0, sd = 1))

Shapiro-Wilk normality test

data:  rnorm(5000, mean = 0, sd = 1)
W = 0.9996, p-value = 0.5105
```

4.5.8.4 An Application to Molecular Biology

In this subsubsection, we consider an application to molecular biology of the toroidal probability models and the methods developed in preceding sections.

On the torus, it is natural to use two angles as coordinates. To describe the so called *dihedral angles* ϕ and θ (sometimes called *conformational angles* or *torsional angles*) in the protein main chain, people use the so called *Ramachandran map*. The *Ramachandran map* identifies a point in the protein main chain with a point on a flat square of the Euclidean plane $\mathbb{R}^2$ with opposite sides identified. From the mathematical point of view, the *Ramachandran map* represents the embedding of the *Flat Torus* (also called *Clifford Torus*) into the Euclidean space $\mathbb{R}^4$. However, it turns out to be a very useful starting point in the inference process.

We consider the data set in [32]. This data set is taken from the open access *Conformational Angles DataBase*. It consists of 8190 *Conformation Angles* from 1208 PDB structures in 25% non-redundant protein chains. See Figure 4.9 and 4.10. The experiment method taken into consideration is the NMR (*Nuclear Magnetic Resonance*).

```
#Data
require(gdata);
DihedralAnglesData = read.xls ("/Users/selvit/Documents/Distribution of
```

```
 Points on Manifolds/Comprehensive Examination
 2015/DehidralAnglesData_JustAngles_CorrectDataGoodFormat.xls",
  sheet = 1, header = TRUE);
DAD<-DihedralAnglesData;
head(DAD);
plot(DAD[,1],DAD[,2]);
plot(DAD[,1],DAD[,2],main="Ramachandran Plot", xlab="phi-angle(degrees)",
ylab="psi-angle(degrees)");
AngoliMedi<-c(mean(DAD[,1]), mean(DAD[,2]));
var(DAD);
smoothScatter(DAD[,1],DAD[,2],main="Ramachandran Plot-SmoothScatter",
xlab="phi-angle(degrees)", ylab="psi-angle(degrees)")
x=sin(2*pi*DAD[,1]/360)/(1+cos(2*pi*DAD[,1]/360));
y=sin(2*pi*DAD[,2]/360)/(1+cos(2*pi*DAD[,2]/360));
Projected<- matrix(c(x,y), nrow = 8190, ncol=2);
dim(Projected);
plot(Projected[,1], Projected[,2], xlim=c(-20,20), ylim=c(-20,20))

#Estimation:
mean(x)
[1] -2.092306
mean(y)
[1] 2.257566
InvProjMeanX<- 2*atan(mean(x))
InvProjMeanY<- 2*atan(mean(y))
InvProjMeanX
[1] -2.249901
InvProjMeanY
[1] 2.307633


#Test
library(mvnormtest)
require(mvnormtest)
mshapiro.test(t(Projected[1:4000,]))
```

```
Shapiro-Wilk normality test
data:  Z
W = 0.1063, p-value < 2.2e-16

#Reject the test of Normality without any doubt,
accordingly to the Shapito Test.
```

The *Ramachandran Plot* for the *residue ALA* is shown below in Figure 4.9 and 4.10. The *Shapiro-Wilk Normality Test* rejects beyond every reasonable doubt the hypothesis of normality on the projected data set and so we can conclude that these data are not distributed as an ISND.

At this point, without taking advantage of the projected data set, it seems not easy to find a reasonable model which fits the data well.

Let us see how the projected marginals behave under *Stereographic Projection*:

```
 par(mfrow=c(2,2)); hist(x,breaks=10000, xlim=c(-10,10));
  hist(y,breaks=10000, xlim=c(-10,10));
```

It seems reasonable to test if $-X \sim exp(\lambda)$ and if $Y \sim exp(\mu)$ for some $\lambda > 0$ and $\mu > 0$. To do this, we perform a *Kolmogorov-Smirnov Test*.

Remark 4.5.53. *Although this hypothesis has been formulated after looking at the data, here we are not interested in the result of the test, but in proposing test procedures. A similar point of view has already been taken in [84] (See Section 5.3 of that paper).*

```
%Maybe the distribution is different. Using the Histogram,
it seems that the projected marginals look like exponentials...
ks.test(y,mean(y))

Two-sample Kolmogorov-Smirnov test

data:  y and mean(y)
D = 0.8172, p-value = 0.5165
alternative hypothesis: two-sided

ks.test(-x,mean(-x))
```

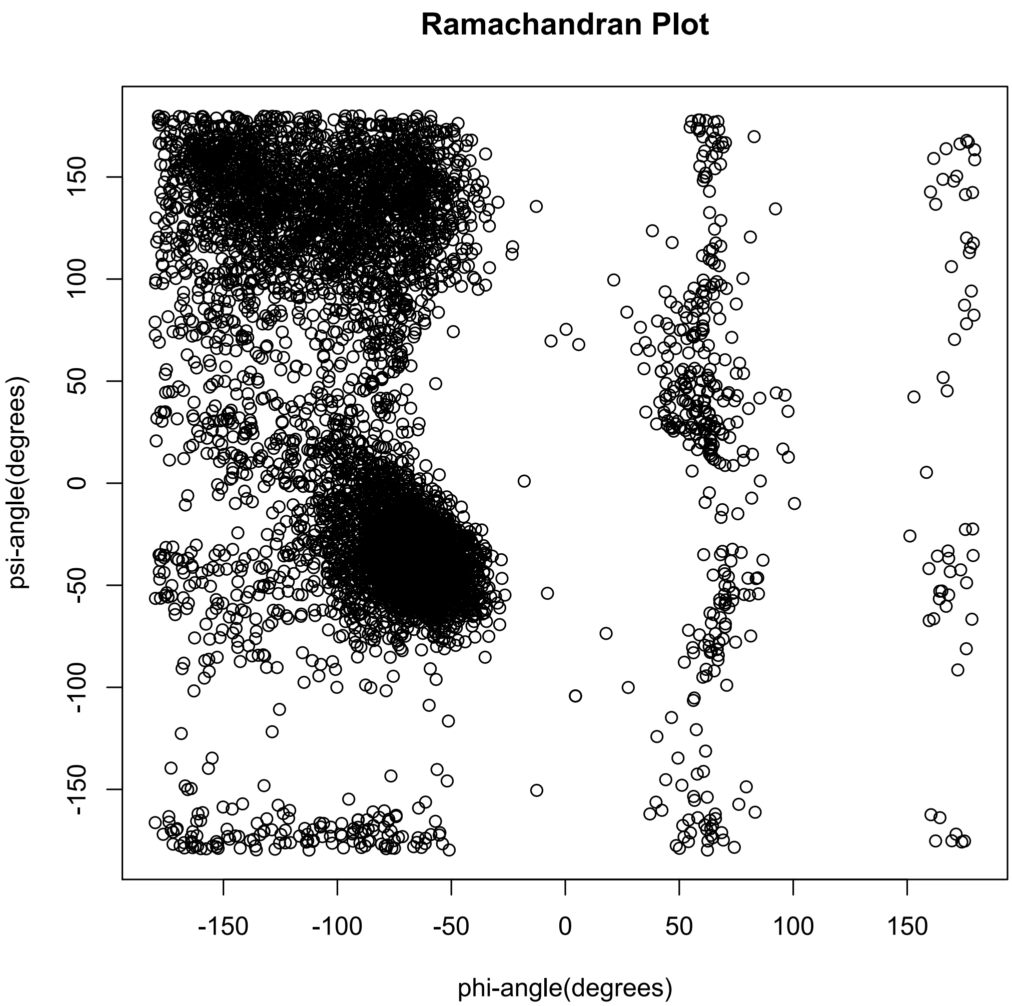


Figure 4.9: Ramachandran Plot of the 8190 *Conformation Angles* from 1208 PDB structures in 25% non-redundant protein chains

```
Two-sample Kolmogorov-Smirnov test

data:  -x and mean(-x)
D = 0.8641, p-value = 0.4442
alternative hypothesis: two-sided
#The Projected look like exponential (the y) and reversed
exponential (the x). This may suggest a stereographic
 double exponential as a good fit. Can't reject, it seems a good fit.
```

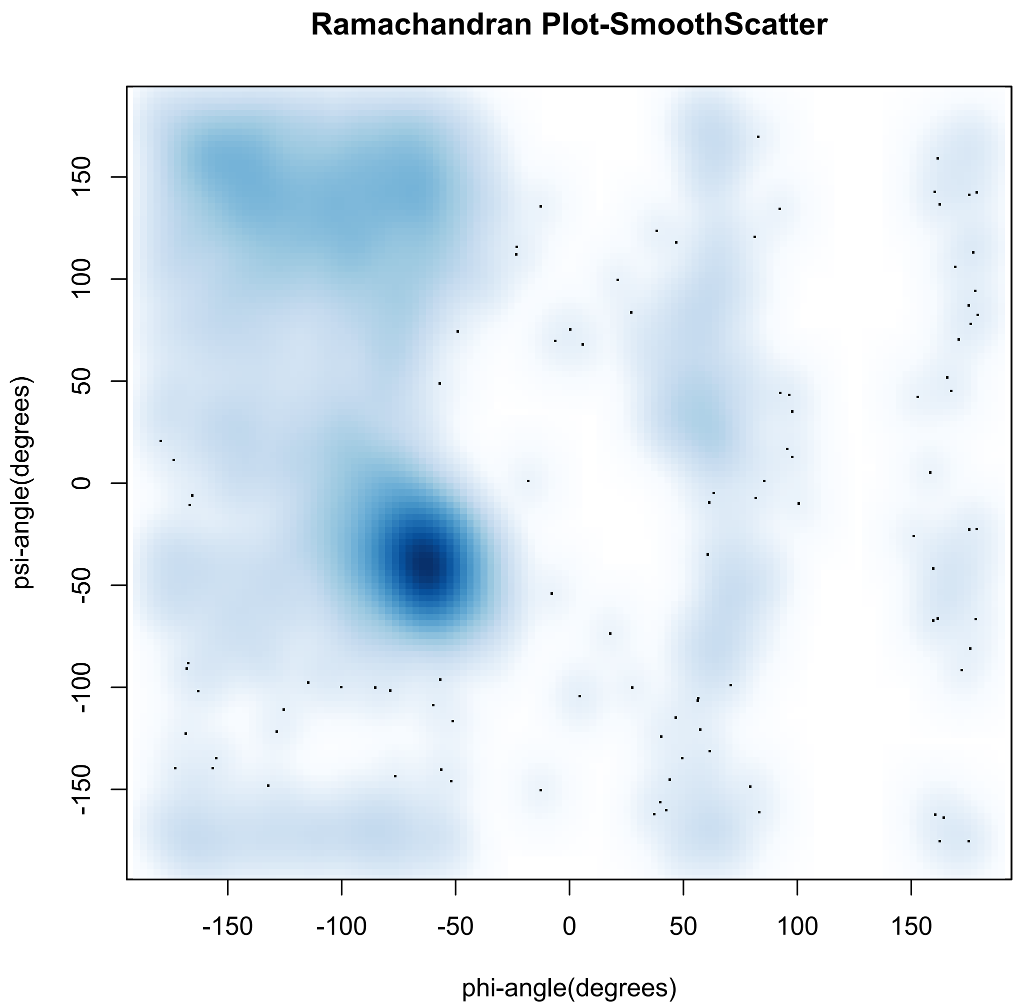


Figure 4.10: Ramachandran Plot of the 8190 *Conformation Angles* from 1208 PDB structures in 25% non-redundant protein chains

From the *Kolmogorov-Smirnov Test*, we cannot reject the *Null Hypothesis* of exponentiality, which does support the assumption of our model.

From the *Ramachandran Plot*, it is clear that there are two peaks and that the data are clustered along two main major circles in the torus. Our model, even if it fits well, does not see these differences.

Remark 4.5.54. *Note that, from this example, it becomes clear how helpful it is to project the data to the Euclidean Space, find a good model there and test for it, and then inverse*

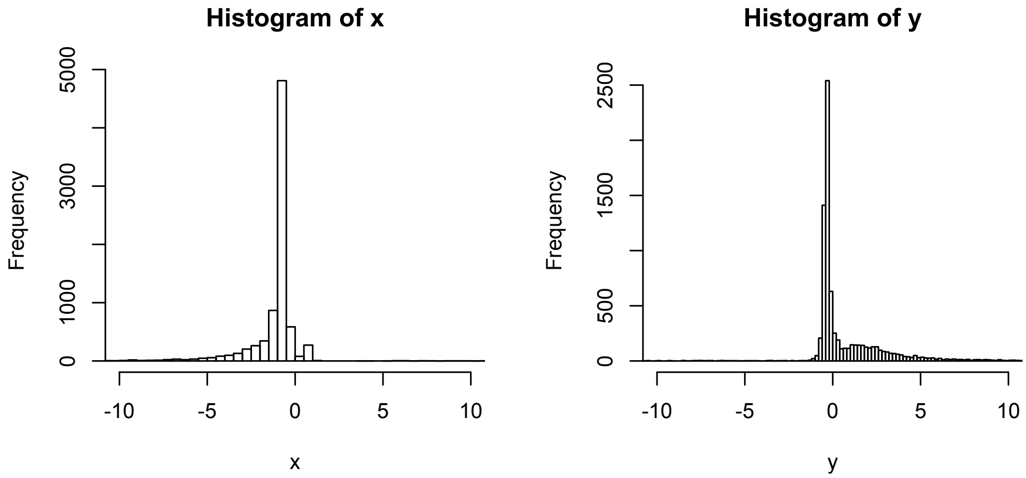


Figure 4.11: Stereographic projection of the marginals of the distributions of the 8190 *Conformation Angles*

project the result to the torus.

4.6 Concluding Remarks

This chapter has dealt with some new results about distributions with symmetries. First, we discussed a result on symmetric order statistics. We proved that the symmetry of any of the order statistics is equivalent to the symmetry of the underlying distribution in both the continuous and discrete case. Note that, even if our result is confined to dimension one, it might be important also in the study of those depth functions which are essentially radial. Second, we characterized elliptical distributions through group invariance and give some properties. A key point here has been the bijective correspondence of the angle/radius coordinates to the Euclidean coordinates. Third, we studied geometric probability distributions on the torus and applied our results to a problem in molecular biology. The new distributions introduced are generated through stereographic projection. We gave several properties of these distributions and compared them with the Von-Mises distribution and its multivariate extensions. The simplicity of this approach is reflected in the simplification of the estimation procedure and numerical algorithms.

Chapter 5

Summary and Future Directions

In this thesis, we have presented some new contributions to distribution theory both for discrete and continuous random variables. The results were connected to their applications.

In chapter 1, we discussed a variational characterization of the Multivariate Normal distribution (MVN), by means of Strichartz Estimates. With respect to the characterization of the MVN distribution as a maximizer of the entropy functional, the characterization as a maximizer of the Strichartz Estimates does not require the constraint of fixed variance. In this chapter, we computed the precise optimal constant for the whole range of Strichartz admissible exponents, discussed the connection of this problem to Restriction Theorems in Fourier analysis and gave some statistical properties of the family of Gaussian Distributions which maximize the Strichartz estimates, such as Fisher Information, Index of Dispersion and Stochastic Ordering. We concluded this chapter presenting an optimization algorithm to compute numerically the maximizers. This last part deserves further development and it is object of current research. Furthermore, Strichartz estimates are available for several dispersive PDEs and there might be characterizations of other probability distributions based on some maximization procedure related to other differential equations. This also deserves further consideration.

Chapter 2 concerned the characterization of distributions using ideas from Optimal Transportation and the Monge-Ampère equation. We gave emphasis to methods to do statistical inference for distributions that do not possess good regularity, decay or integrability properties, like the Cauchy distribution. The main tool used here was a modified version of the characteristic function. We related our results to some topics in Big Data analysis and in particular the Consensus Monte Carlo Algorithm. Further characterizations of probability distributions might be available using these techniques and might be good to work on them in the future.

In chapter 3, we studied the Simpson's Paradox. The Simpson's Paradox is the phenomenon that appears in some datasets, where subgroups with a common trend (say, all negative trend) show the reverse trend when they are aggregated (say, positive trend). We discussed its ubiquity in science with some basic examples from arithmetic, geometry, linear algebra, statistics, game theory, gender bias in university admission and other fields. Our main new results concerned the occurrence of the Simpson's Paradox in Quantum Mechanics. We proved that the Simpson's Paradox occurs for solutions of the Quantum Harmonic Oscillator both in the stationary case and in the non-stationary case. We proved that the phenomenon is not isolated (stability) and that it appears (asymptotically) in the context of the Nonlinear Schrödinger Equation as well. When considering the physical quantities involved with their proper values, from our analysis, it results that the Simpson's Paradox occurs when the Bohr radius is crossed and so our results might have important physical consequences. It might be possible to detect this phenomenon experimentally and so further investigation is needed. Other subjects in this area which deserve further investigations are the probability of occurrence of the paradox and the possible very general algebraic structure that underlines the phenomenon. Note that related problems appear in Spatial Statistics and Meta-Analysis as well.

Chapter 4 contained some new results about distributions with symmetries. First, we discussed a result on symmetric order statistics. We proved that the symmetry of any of the order statistics is equivalent to the symmetry of the underlying distribution in both the continuous and discrete case. Note that even if our result is essentially one dimensional, it might be extended to the multidimensional case, since several important depth functions are radial. Second, we characterized elliptical distributions through group invariance and give some properties. We think that the essence of our characterization is related to the bijective correspondence of the angle/radius coordinates to the Euclidean coordinates and so that extensions, for example, to hyperbolic symmetries or to distributions with star-shaped contours are possible. Third, we studied geometric probability distributions on the torus with applications to molecular biology. We introduced a new way of producing families of distributions generated through stereographic projection and we compared them with the Von-Mises distribution and its multivariate extensions. We added several other properties of these distributions. The simplicity of this approach is reflected in the simplification of the estimation procedure. It seems reasonable to pursue further this point of view and extend it to distributions with more complicated symmetries.

Bibliography

[1] S. I. Amari and H. Nagaoka, Methods of Information Geometry, Methods of information geometry, Translations of Mathematical Monographs, **191**, American Mathematical Society, Providence, RI; Oxford University Press, Oxford, 2000.

[2] B. C. Arnold, N. Balakrishnan and H. N. Nagaraja, *A First Course in Order Statistics*, John Wiley & Sons, New York (1992).

[3] Anonymous, http://www.matrixlab-examples.com/simple-animation.html.

[4] C. Atkinson and A. F. S. Mitchell, Rao's Distance Measure, *Samkhyã-The Indian Journal of Statistics*, **43** (1981) 345-365.

[5] T. Aubin, *Nonlinear analysis on manifolds. Monge-Ampère equations*, Grundlehren der mathematischen Wissenschaften, vol. **252**, Springer-Verlag, Berlin and New York (1982).

[6] T. Aubin, Problèmes isopérimétriques et espaces de Sobolev, *J. Differential Geometry*, **11** no. 4 (1976) 573-598.

[7] D. Ayala and B. Protas On maximum enstrophy growth in a hydrodynamic system *Physica D* 240 (2011) 1553-1563.

[8] L. Baringhaus, Testing for Spherical Symmetry of a Multivariate Distribution, *The Annals of Statistics*, **19** (1991) 899-917.

[9] J. Bennett, N. Bez, C. Jeavons and N. Pattakos, On sharp bilinear Strichartz estimates of Ozawa-Tsutsumi type, *J. Math. Soc. Japan* **69** no. 2 (2017) 459-476.

[10] R. Beran, Testing for Elliptical Symmetry of a Multivariate Density, *The Annals of Statistics*, **7** (1979) 150-162.

[11] H. Berestycki and P. L. Lions, Nonlinear Scalar field equations, *Arch. Rational Mech. Anal.*, **82** no.3 (1983) 313-345.

[12] H. Berestycki, P. L. Lions and L. A. Peletier, An ODE approach to the existence of positive solutions for semilinear problems in $\mathbf{R}^N$, *Indiana Univ. Math. J.* **30** no. 1 (1981) 141-157.

[13] F. A. Berezin and M. Shubin, *The Schrödinger Equation*, Mathematics and its Applications (Soviet Series) **66**, Kluwer Academic Publishers Group, Dordrecht (1991).

[14] D. J. Best and N. I. Fisher, Efficient simulation methods of the Von Mises Distribution, *Appl. Statist.*, **28** (1979) 152-157.

[15] P. J. Bickel, E. A. Hammel and J. W. O'Connell, Sex Bias in Graduate Admissions: Data From Berkeley, *Science*, **187** (4175) (1975) 398-404.

[16] P. Billingsley, Probability and Measure, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ (2012).

[17] C. R. Blyth, On Simpson's Paradox and the Sure-Thing Principle, *Journal of the American Statistical Association*, **67** (338) (1972) 364-366.

[18] J. Bourgain, Refinements of Strichartz' inequality and applications to 2D-NLS with critical nonlinearty, *Int. Math. Res. Notices* **5** (1998) 253-283.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge (2009).

[20] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, *Comm. Pur Appl. Math.*, **44** (1991) no. 4 375-417.

[21] A. Bulut, Maximizers for the Strichartz inequalities for the wave equation, *Differential and Integral Equations*, **23** (2010) 1035-1072.

[22] T. T. Cai, J. Fan and T. Jiang, Distributions of angles in random packing on spheres, *Journal of Machine Learning Research*, **14**, (2013) 1837-1864.

[23] S. Cambanis, S. Huang, S. and G. Simons On the theory of elliptically contoured distributions, *Journal of Multivariate Analysis*, **11**, (1981) 368-385.

[24] E. Carneiro, A Sharp Inequality for the Strichartz Norm, *Int. Math. Res. Not.*, **16** (2009) 3127-3145.

[25] G. Casella, R. L. Berger, Statistical Inference (Second Edition), Duxbury, Pacific Grove, CA (2002).

[26] K.S. Chou, X.J. Wang, Entire solutions of the Monge-Ampère equation, *Comm. Pure Appl. Math.*, **49** (1996) 529-539.

[27] M. Christ and Q. René, Gaussians rarely extremize adjoint Fourier restriction inequalities for paraboloids, *Proc. Amer. Math. Soc.*, **142** (2014) no. 3 887-896.

[28] M. Christ, S. Shuanglin, Existence of extremals for a Fourier restriction inequality, *Anal. PDE*, **5** (2012) no. 2 261-312.

[29] M. Christ, S. Shuanglin, On the extremizers of an adjoint Fourier restriction inequality, *Adv. Math.*, **230** no. 3 (2012) 957-977.

[30] J. P. R. Christiansen, On some measures analogous to Haar Measure, *Math. Scand.*, **26** (1970), 103-106.

[31] S. Cialdi and M. G. A. Paris, The data aggregation problem in quantum hypothesis testing, *Eur. Phys. J. D*, (2015) **69: 7** (4pp).

[32] `http://cluster.physics.iisc.ernet.in/cgibin/cadb/callapplet1.pl?identity=25&experiment=N&mode=F&Entire=1&Fully=0&Additional=0&Generous=0&Disallowed=0&Alpha=0&Beta=0&Threeten=0&residue0=ALA`

[33] K. Conrad, Probability Distributions and Maximum Entropy, *http://www.math.uconn.edu/ kconrad/blurbs/analysis/entropypost.pdf*.

[34] S. I. R. Costa, S. A. Santos and J. E. Strapasson, Fisher information distance: a geometrical reading, *Discrete Applied Mathematics*, Volume **197** (2015) 59-69.

[35] S. I. R. Costa, S. A. Santos and J. E. Strapasson, Fisher Information Matrix and Hyperbolic Geometry, *Proc of IEEE ISOC ITW2005 on coding and complexity*, (2005) 34-36.

[36] H. S. M. Coxeter, Discrete groups generated by reflections, *Ann. of Math.*, **35** (1934) 588-621.

[37] H. S. M. Coxeter, Regular Polytopes (Third Edition), Dover Publications, New York (1973).

[38] H. A. David and H. N. Nagaraja, *Order Statistics*, Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ (2003).

[39] P. Diaconis, *Groups Representations in Probability and Statistics*, Institute of Mathematical Statistics Lecture Notes-Monograph Series **11**, Institute of Mathematical Statistics, Hayward, CA (1988).

[40] P. Diaconis, S. Holmes and M. Shahshahani, Sampling from a Manifold, *Institute of Statistical Science* Volume **10** (2013) 102-125.

[41] D. B. Dix Mathematical Models of Protein Folding, *people.math.sc.edu/dix/fold.pdf*

[42] M. P. DoCarmo *Differential geometry of curves and surfaces* , Prentice-Hall, Inc., Englewood Cliffs, NJ (1976).

[43] B. Dodson, Global well-posedness and scattering for the defocusing, L^2-critical, nonlinear Schrödinger equation when $d = 1$, *Amer. J. Math.*, **138** no. 2 (2016) 531-569.

[44] B. Dodson, Global well-posedness and scattering for the defocusing, L^2-critical, nonlinear Schrödinger equation when $d = 2$, *Duke Math. J.*, **165** no. 18 (2016) 3435-3516.

[45] B. Dodson, Global well-posedness and scattering for the defocusing, L^2-critical, nonlinear Schrödinger equation when $d \geq 3$, *J. Amer. Math. Soc.* **25** no. 2 (2012) 429-463.

[46] T. Duyckaerts, F. Merle, S. Roudenko, Threshold solutions for the focusing 3D cubic Schrödinger equation, *Rev. Mat. Iberoam.* **26** no. 1 (2010) 1-56.

[47] M. L. Eaton, *Group Invariance Applications in Statistics*, NSF-CBMS Regional Conference Series in Probability and Statistics, **1**, Institute of Mathematical Statistics, Hayward, CA; American Statistical Association, Alexandria, VA (1989).

[48] L. Fanelli, L. Vega, N. Visciglia, Existence of maximizers for Sobolev-Strichartz inequalities, *Adv. Math.*, **229** no. 3 (2012) 1912-1923.

[49] L. Fanelli, L. Vega, N. Visciglia, On the existence of maximizers for a family of restriction theorems, *Bull. Lond. Math. Soc.*, **43** no. 4 (2011) 811-817.

[50] K. T. Fang and T. W. Anderson, *Statistical Inference in Elliptically Contoured and Related Distributions*, Allerton Press, New York (1990).

[51] K. T. Fang, S. Kotz and K. W. Ng, Symmetric Multivariate and related Distributions, Monographs on Statistics and Applied Probability, **36**, Chapman and Hall, Ltd., London (1990).

[52] A. Figalli and G. De Philippis, The Monge-Ampère equation and its link to optimal transportation, *Bull. Amer. Math. Soc.*, **51** no. 4 (2014) 527-580.

[53] D. Foschi, Maximizers for the Strichartz inequality, *J. Eur. Math. Soc.*, **9** no. 4 (2007) 739-774.

[54] D. Foschi, Inhomogeneous Strichartz estimates, *J. Hyperbolic Differ. Equ.*, **2** (2005) no. 1 1-24.

[55] M. Ghosh, G. Meeden, On the Non-Attainability of Chebyshev Bounds, *Amer. Statist.* **31** (1977) 35-36.

[56] B. Gidas, W. M. Ni and L. Nirenberg, Symmetry and related properties via the maximum principle, *Comm. Math. Phys.*, **68** (1979) 209-243.

[57] H. H. Goltz and M. L. Smith, Yule-Simpson's Paradox in Research, *Practical Assessment, Research and Evaluation*, **15** Number 15 (2010) 1-9.

[58] E. Gomez, M. A. Gomez-Villegas and J. M. Marin A survey on continuous elliptical vector distributions, *Rev. Mat. Complut.* **16** no. 1 (2003) 345-361 .

[59] I. J. Good and Y. Mittal, "The Amalgamation and Geometry of Two-by-Two Contingency Tables", *The Annals of Statistics*, **15** (2) (1987) 694-711.

[60] M. M. Hingorani and M. O'Donnell, Toroidal proteins: Running rings around DNA, *Current Biology*, Volume **8** Issue 3 (1998) R83-R86.

[61] M. M. Hingorani and M. O'Donnell, A tale of toroids in DNA metabolism, *Nature Reviews Molecular Cell Biology*, **1** (2000) 22-30.

[62] J. E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge University Press, Cambridge, UK (1992).

[63] D. Hundertmark and V. Zharnitsky, On sharp Strichartz inequalities in low dimensions, *Int. Math. Res. Not.* (2006) Art. ID 34080 18 pp.

[64] J. L. Jensen, On the Hyperboloid Distribution, *Scandinavian Journal of Statistics*, Vol. **8** No.4 (1981) 193-206.

[65] J. Jiang and S. Shao, On characterisation of the sharp Strichartz inequality for the Schrödinger equation, *Anal. PDE*, **9** no. 2 (2016) 353-361.

[66] P. E. Jupp and K. V. Mardia, *Directional Statistics*, John Wiley & Sons, New York (2008).

[67] T. Kariya and M. Eaton, Robust Tests for Spherical Symmetry, *The Annals of Statistics*, **5** (1977) 206-215.

[68] M. Keel, T. Tao, Endpoint Strichartz estimates, *Amer. J. Math.*, **120** no. 5 (1998) 955-980.

[69] J. T. Kent, K. V. Mardia and C. C. Taylor, Modelling strategies for bivariate circular data, In: Barber, S., Baxter, P.D., Gusnanto, A., Mardia, K.V. (eds.) *The Art and Science of Statistical Bioinformatics*, Leeds University Press, Leeds (2008) 70-74.

[70] S. Klainerman, Lecture Notes in Analysis (2011), *https://web.math.princeton.edu/ seri/homepage/courses/Analysis2011.pdf.*

[71] S. Klainerman and M. Machedon, Remark on Strichartz-type inequalities, *Int. Math. Res. Not.*, **5** (1996), 201-220.

[72] S. Kotz, N. Balakrishnan and N. L. Johnson, *Continuous Multivariate Distributions*, Vol. 1: Models and Applications (Second Edition), Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York (2000).

[73] M. Kunze, On a variational problem with lack of compactness related to the Strichartz inequality, *Calc. Var. Partial Differential Equations*, **19** no. 3 (2004) 307-336.

[74] M. Kunze, On the existence of a maximizer for the Strichartz inequality, *Comm. Math. Phys.*, **243** no. 1 (2003) 137-162.

[75] M. K. Kwong, Uniqueness of positive solutions of $\Delta u - u + u^p = 0$ in $\mathbf{R}^n$, *Arch. Rational Mech. Anal.*, **105** (1989) 243-366.

[76] E. L. Lehmann, Ordered families of distributions, *The Annals of Mathematical Statistics*, **26** (1955) 399-419.

[77] M. Lovrić, M. Min-Oo and E. A. Ruh, Multivariate normal distributions parametrized as a Riemannian symmetric space, *Journal of Multivariate Analysis*, **74** (2000) 36-48.

[78] G. Malinas and J. Bigelow, Simpson's Paradox, *The Stanford Encyclopedia of Philosophy* (2012), Edward N. Zalta, *http://plato.stanford.edu/archives/win2012/entries/paradox-simpson/.*

[79] K. V. Mardia, Statistics of Directional Data (with discussion), *J. R. Stat. Soc. B* (Methodology) **37** (1975) 349-393.

[80] K. V. Mardia, Bayesian Analysis for Bivariate Von Mises Distributions, *J. Appl. Stat.*, **37** (2010) 515-528.

[81] K. V. Mardia, Characterization of directional distributions, In: Patil, G.P., Kotz, S., Ord, J.K. (eds.) *Statistical Distributions in Scientific Work*, D. Reidel Publishing Company, Dordrecht/Holland (1975) 365-386.

[82] K. V. Mardia and S. A. M. El-Atoum, Bayesian inference for the von Mises-Fisher distribution, *Biometrika*, **63** (1976) 203-206.

[83] K. V. Mardia and J. Frellsen, Statistics of bivariate von Mises distributions, *Bayesian methods in structural bioinformatics*, Stat. Biol. Health, Springer, Heidelberg (2012) 159-178.

[84] K. V. Mardia, J. T. Kent and G. Hughes, C. C. Taylor and H. Singh, A multivariate von Mises distribution with applications to bioinformatics, *Can. J. Stat.*, **36** (2008) 99-109. (2012) 159-178.

[85] K. V. Mardia, J. T. Kent and G. Hughes and C. C. Taylor, Maximum likelihood estimation using composite likelihoods for closed exponential families, *Biometrika*, **96** (2009) 975-982.

[86] K. V. Mardia and R. J.Marshall, Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, **71** (1) (1984) 135-146.

[87] K. V. Mardia, C. C. Taylor and G. K. Subramaniam, Protein Bioinformatics and Mixtures of bivariate von Mises distrbutions for angular data, *Biometrics*, **63** (2007) 505-512.

[88] Y. Martel and F. Merle, Multi solitary waves for the nonlinear Schrödinger equations, *Ann.I.H. Poincaré*, **23** (2006) 849-864.

[89] P. McCullagh, Möbius Transformation and Cauchy Parameter Estimation, *Ann. Stat.*, **24** no.2 787-808.

[90] R. J. Muirhead, *Aspects of multivariate statistical theory*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York (1982).

[91] F. Norris, Can Every Group Be Worse Than Average ? Yes.

economix.blogs.nytimes.com/2013/05/01/
can-every-group-be-worse-than-average-yes/

[92] S. O'Hagan Uniform Sampling Methods for Various Compact Spaces, Msc Thesis McMaster University (2007).

[93] M. G. A. Paris, Two quantum Simpson's paradoxes, *J. Phys. A: Math. Theor.*, **45** (2012) (7pp).

[94] M. G. Pavlides and M. D. Perlman, How likely is Simpson's Paradox?, *The American Statistician*, **63** (2009), 226-233.

[95] K. Pearson, A. Lee and L. Bramley-Moore, Genetic (reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred racehorses, *Philosophical Transactions of the Royal Society A*, **192** (1899) 257-330.

[96] A. D. Polyanin and V. F. Zaitsev, *Handbook of Nonlinear Partial Differential Equations*, Chapman and Hall/CRC, Boca Raton, FL (2004).

[97] G. N. Ramachandran, V. Ramakrishnan and C. Sasisekharan, Stereochemistry of polypeptide chain configurations *J. Mol. Biol.*, **7** (1963) 95-99.

[98] C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, *Bullettin of the Calcutta Math. Soc.*, **37** (1945) 81-91.

[99] L. Rivest, A distribution for dependent unit vectors, *Comm. Stat. Theor. Meth.* **17**, 461-483. (1988)

[100] P. Royston, An extension of Shapiro and Wilk's W test for normality to large samples, *Applied Statistics*, **31** (1982) 115-124.

[101] P. Royston, Algorithm AS 181: The W test for Normality, *Applied Statistics*, **31** (1982) 176-180.

[102] P. Royston, Remark AS R94: A remark on Algorithm AS 181: The W test for normality, *Applied Statistics*, **44** (1995) 547-551.

[103] J. P. Royston, Some techniques for assessing Multivariate Normality Based on the Shapiro-Wilk's W, *Applied Statistics*, **32** (1983) 121-133.

[104] W. Rudin, *Real and Complex Analysis* (Third edition), McGraw-Hill Book Co., New York (1987).

[105] S. L. Scott , A. W. Blocker, F. V. Bonassi, H. A. Chipman, E. I. George, and R. E. McCulloch, *Bayes and Big Data: The Consensus Monte Carlo Algorithm*, http://research.google.com/pubs/pub41849.html

[106] D. J. Selkoe, Folding proteins in fatal ways, *Nature*, **426** (6968) (2003) 9004.

[107] S. S. Shapiro and M. B. Wilk, Analysis of variance test for normality (complete samples), *Biometrika*, **52** (1965) 591-611.

[108] Y. Shi, Quantum Simpson's Paradox and High Order Bell-Tsileron Inequalities, preprint available at *arxiv.org/pdf/1203.2675*.

[109] E. H. Simpson, The Interpretation of Interaction in Contingency Tables, *Journal of the Royal Statistical Society, Series B* (Methodology), **13** (1951) 238-241.

[110] H. Singh, V. Hnizdo and E. Demchuk, Probabilistic Model for two dependent circular variables, *Biometrika*, **89** 3 (2002) 719-723.

[111] D. Stansbury, Sampling From the Normal Distribution Using the Box-Muller Transform,

`https://theclevermachine.wordpress.com/2012/09/11/sampling-from-the-normal-distribution-using-the-box-muller-transform/`

[112] D. Stoyan and W.S. Kendall, *Stochastic Geometry and its Applications* (Third edition), Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester (2013).

[113] R. S. Strichartz, Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations, *Duke Math. J.*, **44** (1977) no. 3 705-714.

[114] R. J. Taggart, Inhomogeneous Strichartz estimates, *Forum Math.*, **22** no. 5 (2010) 825-853.

[115] G. Talenti, Best constant in Sobolev inequality, *Ann. Mat. Pura Appl. (4)* **110** (1976) 353-372.

[116] T.Tao, *Nonlinear Dispersive Equations: Local and Global Analysis*, BMS Regional Conference Series in Mathematics, Number **106**, Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI (2006).

[117] T. Tao, Some recent progress on the restriction conjecture, Fourier analysis and convexity, *Appl. Numer. Harmon. Anal*, Birkhäuser Boston, Boston, MA (2004) 217-243.

[118] C. C. Taylor, Directional Data on the torus, with applications to protein structure, *Proceedings of the SIS 2009 Statistical Conference on Statistical Methods for the Analysis of Large Data-Sets*, (2013) 105-108.

[119] Y. Tsutsumi, L^2 solutions for nonlinear Schrödinger equations and nonlinear groups, *Funk. Ekva.*, **30** (1987) 115-125.

[120] C. Villani, *Optimal Transport, old and new*, Fundamental Principles of Mathematical Sciences **338**, Springer-Verlag, Berlin (2009).

[121] C. H. Wagner, Simpson's Paradox in Real Life, *The American Statistician*, **36** (1) (1982) 46-48.

[122] M. Weinstein, Nonlinear Schrödinger equations and sharp interpolation estimates, *Comm. Math. Phys.,* Volume **87** Number 4 (1982) 567-576.

[123] H. Weyl, *The Classical Groups*, Princeton University Press, Princeton, NJ (1973).

[124] `https://en.wikipedia.org/wiki/Reflection_group`

[125] G. U. Yule, Notes on the Theory of Association of Attributes in Statistics, *Biometrika*, **2** (2) (1903) 121-134.

[126] G. U. Yule, *An Introduction to the Theory of Statistics* (Sixth Edition), Griffin, London (1922).

[127] G. U. Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, An Introduction to the Theory of Statistics (Fourteenth Edition), Griffin, London (1950).