Three essays on modeling the dependence between financial assets

Damien Bosc

THÈSE

Pour l'obtention du grade de
Docteur de l'École Polytechnique
Spécialité : Sciences économiques

présentée et soutenue publiquement par

Damien BOSC

à l'École Polytechnique, le 21 juin 2012

Trois essais sur la modélisation de la dépendance entre actifs financiers

---

Three essays on modeling the dependence between financial assets

Directeur de thèse : **Alfred GALICHON**

Membres du jury

Président	:	**Peter**	**TANKOV**	Université Paris-Diderot
Rapporteurs	:	**Guillaume**	**CARLIER**	Université Paris Dauphine
		Marc	**HENRY**	Université de Montréal
Examinateurs	:	**Edouard**	**OUDET**	Université Joseph Fourier
	:	**Eduardo**	**PEREZ**	Ecole Polytechnique
Directeur de thèse	:	**Alfred**	**GALICHON**	Ecole Polytechnique

L' École Polytechnique n'entend donner aucune approbation,
ni improbation aux opinions émises dans les thèses.
Ces opinions doivent êtres considérées comme propres à leur auteur.

Cette thèse a été réalisée dans le cadre d'un contrat CIFRE, au sein d'AXA Investment Managers et du département d'Économie de l'École Polytechnique. Elle a bénéficié du support du Fonds AXA pour la Recherche et de l'Association Nationale pour la Recherche et la Technologie.

Remerciements

Ces travaux ont été réalisés dans le cadre d'une thèse CIFRE au sein d'AXA Investment Managers, et n'auraient pu se faire sans l'appui du Fonds AXA pour la Recherche et de l'Association Nationale de la Recherche et de la Technologie, que je remercie vivement. Je remercie plus particulièrement l'équipe 'Quant' de la plateforme Investment Solutions dans laquelle j'ai travaillé depuis mon stage de fin de Master en 2008. J'ai beaucoup appris au contact des quants, Martin, Ethan, Mohamed, Fabrice, Ichem, Camille, qu'ils en soient ici remerciés. Je tiens en particulier à exprimer ma gratitude à Martin Deux, qui a encadré mon activité dans cette équipe, et qui s'est montré remarquablement disponible, enjoué et peu avare de son expérience de quant qui m'a singulièrement profité.

Je remercie sincèrement mon directeur de thèse, Alfred Galichon, pour sa grande disponibilité dont j'ai amplement bénéficié, son exigence scientifique et sa patience à m'apprendre le métier de chercheur. Cela a été un grand honneur pour moi de faire mes premiers pas dans le monde de la recherche sous sa direction.

Je remercie vivement les rapporteurs Guillaume Carlier et Marc Henry pour leur lecture très attentive de mes travaux et leurs remarques détaillées. Merci à Peter Tankov d'avoir d'accepté la présidence du jury ainsi qu'à Edouard Oudet et Eduardo Perez pour m'avoir fait l'honneur de participer à mon jury de thèse.

Merci à mes coburaux du département d'Économie de l'École Polytechnique, Arnaud, Idrissa, Mathias et aux autres 'collègues' Jean-Philippe, Julie, Sabine, Clémence, Rafael, Antonin. Merci aux assistantes Eliane et Lyza pour leur patience à me fournir l'aide administrative nécessaire.

Mes remerciements également à Christine Ferré de l'Ecole Doctorale d'avoir répondu rapidement à mes nombreux mails et coup de téléphones, forcément urgents, à quelques jours de la soutenance de thèse. Merci à Marie-Claire Leterrier pour son suivi attentif tout au long de mon parcours chez AXA-IM.

Un grand merci à tous les 'copains'. Les copains de lycée Vincent et Baptiste, et ceux qui m'ont soutenu par des petits mots sympathiques : Aliona, Xavier, qui a suivi tout cela de fort près, Julie et Guillaume. Mention pour Cyril qui faisait son doctorat à quelques pas de là au Centre de Physique Théorique de l'École, et avec qui j'ai eu l'occasion d'échanger sur les bonheurs (rares) et les vicissitudes (moins rares) de la vie de thésard.

Enfin merci mes sœurs, à mes parents et à mes beaux-parents, qui m'ont offert les meilleures conditions de travail grâce aux week-ends à Vannes, Tours et La Rochelle. *Last but (definitively) not least*, ma femme Mathilde, qui m'a apporté un soutien sans faille et précieux tout au long de ces dernières années.

Contents

Introduction

Modeling dependence between assets is a subject of crucial importance in finance. It has emerged during the 90's and in the last decade as a necessary improvement of previous models in order to develop and value complex financial products or strategies on several underlyings, whose value is strongly sensitive to the dependence between these underlyings. For instance, basket options, and their sophistications, such as the so-called mountain range options. More generally, many options and strategies require hybrid models to be priced. These are models where two or more class of assets are coupled in order to capture their multivariate behavior and not only their individual dynamics, such as, for instance, equity and interest rates (hybrid Hull-white), equity and volatility (stochastic volatility models). Furthermore, as 'correlation between assets' becomes an asset class, just as volatility became an asset class when investors started to have views on volatility and implemented them through the purchase of derivatives on volatility such as variance swaps, correlation products such as correlation swaps asks for accurate dependence modeling. However, as proven by the 2007 financial crisis, standard models of dependence might be insufficient when the market switch to an extreme regime, and there is clearly a room for improvement for financial models to better represent the assets comovements.

The notion of linear correlation is used as an ubiquitous measure of multivariate dependence. This notion dates back to Francis Galton and Karl Pearson and provides a simple means to quantify the strength of the dependence between two real random variables. It characterizes completely the dependence in Gaussian models. This simplicity explains the presence of correlation in financial models; it stems naturally from the fact that the classic financials models are based on Gaussian distribution. Such models are Markowitz' modern portfolio theory, and the derived Capital Asset Pricing Model, or factors model such as Arbitrage Pricing Theory. In continuous-time finance, a wide class of models is the class of diffusion models with Gaussian noise and, once again, the dependence between assets is often modeled by deterministic correlation parameters or correlation matrices, understating that the dependence is Gaussian. However, correlation soon finds its limits when the marginal distributions are not Gaussian and is not a satisfactory tool when one wants to introduce non Gaussian dependence, see Embrecht [3] for more examples where correlation fails.

The assumption that financial assets have Gaussian dependence is quite restrictive, especially when derivatives products are involved, and this assumption might generate unwanted features and misunderstandings. For all these reasons, the notion of *copula* sparked a vivid interest in the financial community in the last decade. This object has been long known to statisticians and has been used, although unnamed, by Hoeffding and Fréchet in the 40's and 50's. It was well later on used to build financial models of dependence, for instance in credit derivative modeling. A copula is a function which embeds all the possible information on the dependence of several random variables. Hence, modeling the dependence between n assets $X_1, \ldots, X_n$ amounts to fix a copula, which in this case is a function of n variables defined on the unit hypercube $[0,1]^n$. It allows for a clear separation between the information on the marginals, summarized by the individual cumulative distribution functions, and the specific information on the dependence.

Let us give some concrete examples of how copulas have been used in finance. In credit derivatives modeling, a critical feature of a model is to adequately represent the arrival of default times of government bonds, corporate bonds, etc... Li's method [7], which introduced the use of Gaussian copulas in financial modeling, consists in choosing the marginal distributions of each default time, and then choosing a Gaussian copula with some correlation matrix Σ to model the dependence between the default times. Thus, although the distributions of the default times are not Gaussian at all, the dependence between them is the same as the dependence of a Gaussian vector with correlation Σ. This works for any copula and allows to impose any form on dependence; this approach is used in the so-called semi-dynamic copula modeling (see Schönbucher's book [11], p. 337 *et seq*). In equity derivatives pricing, copulas have been used in the same fashion. The simplest case where they have been used is for a European option which pays at maturity T a payoff $g(S_T^1, \cdots, S_T^N)$, where S_T^i the value of underlying i at T. The price of such an option depends on the multivariate distribution of the assets at the maturity, which can be split once again into the marginal distributions and the copula.

These examples give an opportunity to point out two drawbacks of using copulas. Consider the previous example of default times modeling. Assume that we want to value a complex credit derivative (such as a Credit Default Obligation), whose value depend on a pool of assets that can be split into two sets, e.g. corporate bonds on the one hand, and home loans on the other hand. Assume that the two copulas ruling the dependence among each of these set of assets is fixed. The dependence structure of the model is completely specified as soon as the dependence between the two sets is defined. As we have chosen to model dependence with copulas, it appears natural to define the copula which rules the overall dependence as

$$C_3(C_1(u_1, \ldots, u_N), C_2(v_1, \ldots, v_M))$$

where C_1 is the copula for the N corporate bonds default times, C_2 is the copula for the M home loans default times and C_3 is a bivariate copula that handles the dependence between the two. However, such a formula does not define a copula in general, unless C_3 is the independence copula. This results from an impossibility theorem proved by Genest et al. [4]. More generally, it is a stylized fact of copula theory that defining a multivariate copula is difficult. As a result, the copula might not be the tool to be preferred when facing the problem of aggregating the dependence.

A second drawback of copula functions is their static nature. Indeed, while they are well suited to the valuation of financial derivatives whose prices depend on the distribution of the assets at a single time, they are more difficult to use in a dynamic framework. For instance, a desirable feature of a pricing model is to give the assets a Markovian dynamics. Combining this Markovian feature and the modeling of the 'spatial dependence', i.e. the cross-sectional dependence between the assets, with copulas is no easy task. There exists an abundant literature on copulas and time-series (see Patton [9] and references therein), where the copulas are used in a (discrete) dynamic setting. Furthermore, the time dependence structure of (possibly continuous) real-valued Markov processes is well understood in terms of copulas, see Darsow et al. [1] and Ibragimov [5]. However, to the best of our knowledge, the problem of modeling the dynamic spatial dependence of continuous Markov processes by copulas has not been thoroughly solved yet.

More generally, copulas are defined from a scaling of the marginals by their cumulative distribution function. While this is natural for a univariate random variable, as the resulting variable has a uniform distribution over $[0,1]$, this does not make sense *a priori* for a multivariate distribution, as well as the notion of quantile. Nevertheless, the problem of scaling one distribution to another, i.e. finding a deterministic function $f_{\mu,\nu}$ such that $f_{\mu,\nu}(X)$ has law ν if X has law μ, still remains. Similarly, multivariate dependence problems can be formulated in the same manner as their univariate analogs, such as determining which distribution maximizes correlation when the marginals distributions are fixed:

$$\sup_{\substack{X\sim p\\ Y\sim q}} \mathbf{E}(XY)$$

If μ and ν are probability distributions over $\mathbf{R}$ this problem amounts to find the copula which maximizes the correlation between the marginals, but it does still perfectly make sense if μ and ν are probabilities over $\mathbf{R}^N$, and the product is replaced by a scalar product:

$$\sup_{\substack{X\sim p\\ Y\sim q}} \mathbf{E}(X\cdot Y)$$

This sort of *fixed marginals problems* in the context of *multivariate* dependence cannot be

tackled by the same tools as in the univariate case.

This thesis aims at addressing the two topics sketched above, namely dependence aggregation and dependence modeling for Markovian diffusions. The two first chapters tackle the issue of multivariate dependence, i.e. dependence between multivariate marginal distributions. More precisely, these chapters aim at describing, characterizing and computing extreme multivariate dependence between random vectors. The third chapter shows how copulas can be used to model the spatial dependence between two Markovian diffusions, and is a first step to build genuine continuous-time models of dependence with copulas.

Chapter 1 The first chapter studies the modeling of the dependence between random vectors with fixed marginals, and in particular the notion of comonotonicity between random vectors. Recall that two real random variables X and Y are comonotone if and only if they can be written as an increasing function of a third variable. A simple procedure to obtain a pair of comonotonic variables (X, Y) with marginal distributions μ and ν is to consider $(X, F_\nu^{-1} \circ F_\mu(X))$ where F_μ is the cumulative distribution of μ, F_ν^{-1} the quantile function of ν and X is a random variable with law μ. Thus, a deterministic scaling of one distribution to another is obtained, and this transform has moreover a particularly simple expression. However this approach fails if the marginals are not univariate. The complexity of the multivariate case can be seen from the fact that if μ is a probability on $\mathbf{R}^N$, the law of the variable $F_\mu(X)$, $X \sim \mu$, is not the uniform law on the unit hypercube, contrary to the one dimensional case, and there is no unique notion of multivariate quantile. Nevertheless, optimal transport theory proposes a scaling of one multidimensional law of probability to another and therefore proposes a possible definition of multivariate comonoticity.

More precisely, if μ and ν are two probability distributions over $\mathbf{R}^N$, consider the set of probability distributions π over $\mathbf{R}^N \times \mathbf{R}^N$ such that the distribution of the first N-dimensional marginal is μ and the distribution of the second one is ν, i.e. $\pi(A \times \mathbf{R}^N) = \mu(A)$ and $\pi(\mathbf{R}^N \times A) = \nu(A)$ for every Borel set $A \subset \mathbf{R}^N$, and denote $\Pi(\mu, \nu)$ this set. Among these couplings π, a distribution of special interest is the one solving the variational problem

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbf{R}^N \times \mathbf{R}^N} |x - y|^2 d\pi(x, y) \tag{1}$$

This problem has a unique solution and, by definition, this solution minimizes the quadratic distance between the first N-dimensional marginal and the second one. The study of such variational problems is the subject of optimal transport theory and has found applications in many fields, pure mathematics, economics, numerical optimization, medical imaging etc. . . see the book by C. Villani [12] for an introduction to this theory. According to the Monge-Kantorovitch duality, the linear problem (1) admits a dual problem which writes (up to addi-

tive constants):

$$\inf_{\varphi \in L^1(\mu)} \int_{\mathbf{R}^N} \varphi(x) d\mu(x) + \int_{\mathbf{R}^N} \varphi^*(y) d\nu(y) \tag{2}$$

where $\varphi^*(y) = \sup_{x \in \mathbf{R}^N}(x \cdot y - \varphi(x))$ is the Legendre transform of φ. From this duality, it can be shown that the optimal coupling in (1) takes the form (under regularity assumptions on the distributions μ and ν): $\pi_{opt} = (Id \times \nabla\varphi)\#\mu$, where $\nabla\varphi$ is the μ-a.e. unique gradient of a convex function such that $\nabla\varphi\#\mu = \nu$. In other words, the optimal coupling is the law of a pair $(X, \nabla\varphi(X))$ where X has law μ, and, up to an additive function in x and y, φ is solution of the dual problem (2). Considering that $\nabla\varphi$ is somehow the multivariate analog of an increasing function, the optimal coupling exhibits a strong dependence between its marginals, and can be seen as a generalization of the notion of comonotonicity in the multivariate case. This coupling can be used in practice to define multivariate and invariant in law risk measures, such as the maximum correlation: $\rho_Y(X) = \sup_{\substack{\tilde{X} \sim X \\ \tilde{Y} \sim Y}} \mathbf{E}(\tilde{X} \cdot \tilde{Y})$, see Rüschendorf [8].

Unlike the comonotonic coupling $Y = F_\nu^{-1} \circ F_\mu(X)$ in the one dimensional case, which is readily computed, there is in general no analytic formula for the function $\nabla\varphi$. To address the problem of the computation of $\nabla\varphi$, we first treat the case of a discrete target distribution ν. Writing y_i the atoms of ν and $q_i = \nu(\{y_i\})$, the solution of the dual problem is easily seen to be a piecewise affine function $\varphi_v(x) = \max_i(x \cdot y_i - v_i)$, and the dual problem becomes a finite-dimensional variational problem

$$\inf_{v \in \mathbf{R}^N} \int_{\mathbf{R}^N} \varphi_v(x) d\mu(x) + q \cdot v \tag{3}$$

An essential feature of this problem is the convexity and the boundedness of the objective function. Thus problem (3) can be solved by classic techniques, such as gradient descent algorithms. A steepest descent algorithm would read

$$v^{i+1} = v^i - \nabla\mathcal{F}(v^i) \tag{4}$$

where $\mathcal{F}(v)$ is the objective function of (3). Furthermore, this algorithm can be interpreted as a Walrasian auction algorithm, where a finite set of sellers (located at the points y_j) offer a good with supply q_j. The steepest descent (4) mimics the behavior of buyers competing for this good in such a way that the prices adjust so that supply and demand match. An equivalent interpretation is that the primal problem (1) is the social planner's objective — maximizing the total economic surplus — and is equivalent to the dual problem (3) (adjusting supply and demand), which is one of Walras' theorem.

The second part of the chapter consists in showing that discretizing the target measure is a valid approach to approximate an optimal Kantorovitch potential φ. We give a rigorous

statement of the fact that, provided a sequence of discrete measures with a finite number of atoms ν_N converges in law to ν, the sequence of optimal potentials φ_N (which solves problem (3) with target measure ν_N) converges uniformly on the support of the initial measure μ to an optimal potential φ.

Eventually, the last part of the chapter aims at comparing various algorithms that solve the transport problem. More precisely we detail: Bertsekas' auction algorithm, which solves the dual problem (3) when both measures are discrete and equally-weighted, thanks to a repeated auction process; linear programming algorithms that use a slightly different form of the dual problem, namely:

$$\sup_{\substack{\varphi,\psi\in\ \mathbf{R}^N \\ \varphi_i+\psi_j\geq x_i\cdot y_j,\forall i,j}} \varphi\cdot p+\psi\cdot q$$

where μ is assumed discrete with atoms $x_i\in\ \mathbf{R}^N$ and p is the vector of the mass of the atoms of μ; the Iterative Proportional Fitting Procedure which consists in relaxing the primal problem (1) by the addition of an entropy term and then solving the relaxed problem with an analog of Von Neumann's alternative projection algorithm and, finally, quasi-Newton method applied to problem 3. Choosing the classical Bertsekas algorithm as a benchmark, we tested these algorithms on three simple test cases, for which the analytic form of the optimal transport map is known. In these three cases, our experiments show that the combination of the IPFP algorithm that provide a 'warm point', and then the use of a quasi-Newton algorithm beats the other algorithms. These latter combination has an estimated complexity $O(N^{5/2})$, while the numerical speed of convergence is $O(1/\sqrt{N})$.

Chapter 2 While the first chapter aimed at computing the quadratic optimal coupling, also called the maximum correlation coupling, the second chapter aims at providing a simple and wider notion of extreme dependence between random vectors. The maximum correlation coupling, which can be seen as a multivariate comonotonicity coupling, is a rather restrictive model for such dependence, as it only takes component-wise covariances into account. A simple way to define extremal couplings consists in considering the possibility of cross-dependence, and, with the same notations as above, studying the following variational problem

$$\sup_{\pi\in\Pi(\mu,\nu)}\int_{\mathbf{R}^N\times\mathbf{R}^N} x\cdot My\ d\pi(x,y) \tag{5}$$

where M is a given N dimensional square matrix (w.l.o.g. as the case of marginals with different sizes is similar). The maximum correlation coupling corresponds to $M=Id$. Then, according to optimal transport theory (and up to some conditions on the marginal distributions), an optimal coupling solving (5) takes the form $MY=\nabla\varphi_M(X)$, for some convex function φ_M. Such a coupling exhibits the same comonotonicity property as the maximum

correlation coupling, up to a linear transform of one of the variables, and is proposed as a possible definition of multivariate extreme dependence. Such a definition takes into account the cross-dependence of the components of each marginal law, and admits a geometric characterization. Indeed, the extremality of a coupling (X, Y) can be checked on its cross-covariance matrix $\mathbf{E}(XY')$. Introduce the *covariogram*

$$\mathcal{F}(\mu, \nu) = \{\mathbf{E}_\pi(XY'), \pi \in \Pi(\mu, \nu)\}$$

which is the set of all cross-covariance matrices corresponding to couplings whose first N-dimensional marginal has law μ and the second one has law ν. The first part of this chapter proposes the following geometric characterization: the extremal couplings, as defined above, are the couplings $(X, Y) \in \Pi(\mu, \nu)$ such that the cross-covariance matrix $\mathbf{E}(XY')$ is located on the boundary of $\mathcal{F}(\mu, \nu)$. The covariogram is also useful to study another notion of extremality. Namely, consider some convex order $\succ$ on the set $\mathcal{F}(\mu, \nu)$: then the couplings whose cross-covariance matrix is maximal with respect to this order should exhibit some strong form dependence. Actually, with the help of a saddlepoint theorem, one can show that such couplings are also extremal, in the sense that there exists a matrix M belonging to a set of matrix $S_\succ$, such that the coupling solves (5). These couplings are called positive extreme couplings and form a subset of extreme couplings. For instance if $\succ$ is the (strict) Loewner order on matrices, defined by $M \succ N$ iff $M - N$ has a strictly positive symmetric part, then positive extreme couplings are the one maximizing (5) for some nontrivial matrix M whose symmetric part is nonnegative.

In a second part of this chapter, an algorithm is proposed to compute these extreme couplings. Given any coupling $\hat{\pi} \in \Pi(\mu, \nu)$, we would like to find a means of associating $\hat{\pi}$ to an extreme dependent coupling. Geometrically speaking, this amounts to consider a matrix inside the covariogram, and project it on the boundary of the covariogram. Of course, there are several way to perform such a projection, and we propose one which respects the structure of the initial problem (5) and allows for explicit computations. As in the first chapter, an entropic relaxation is used:

$$W_T(M) := \sup_{\pi \in \Pi(\mu,\nu)} \mathbf{E}_\pi(X'MY) + TEnt(\pi) \tag{6}$$

where $Ent(\pi)$ is the entropy of the coupling π, formally $-\mathbf{E}_\pi(\log(\pi(X, Y)))$. This is a perturbed version of the original problem, which can be formulated as a projection problem with respect to the Kullblack-Leibler pseudo-distance. Moreover, if $\sigma_{\hat{\pi}}$ is the cross-covariance matrix of the coupling $\hat{\pi}$, then the first order condition of the following problem

$$\inf_{M \in M_N(\mathbf{R})} W_T(M) - \sigma_{\hat{\pi}} \cdot M \tag{7}$$

reads $\sigma_{\hat{\pi}} = \mathbf{E}_{\pi(T,M)}(XY')$ where $\pi(T, M)$ is a coupling achieving the supremum (6). Therefore,

a means to associate an extremal coupling to $\hat{\pi}$ is to fix T, e.g. $T = 1$, and then find the matrix $\hat{M}$ solving (7). $\hat{M}$ is then the cross-covariance matrix of the coupling $\pi_{1,\hat{M}}$ maximizing (6) with $T = 1$. Thus a whole trajectory of nondeterministic couplings $\pi_{T,\hat{M}}$ obtains that satisfy $\mathbf{E}_{\pi_{1,\hat{M}}}(XY') = \sigma_{\hat{\pi}}$ and $\pi_{0,\hat{M}}$ is an extremal coupling. Moreover, the projection of $\sigma_{\hat{\pi}}$ onto the boundary of the covariogram is particularly simple to compute, at least when the marginal distributions are discrete, as the Iterative Proportional Fitting algorithm is then particularly efficient to solve (6) whilst (7) is a standard convex problem.

The third part of the chapter focuses on applications. First, performing a singular value decomposition of the matrix $\hat{M}$, we exhibit a linear transform of the marginal distributions of an empirical coupling, which allows interpret the extreme coupling $\pi_{0,\hat{M}}$ as the maximum correlation coupling once the marginal distributions are linearly transformed. This linear transform is helpful in dependence problems where two economies are involved to define new indices from financial indices that would solve the problem of maximal correlation under the law of the extreme coupling. Then, we apply this technique to multivariate stress testing: a Markowitz allocation model is considered, and the impact of the change of the dependence between two subsets of the investment universe is assessed. Interestingly, it shows that while the maximum correlation coupling might fail at stressing the portfolio, it is not the case with the previous method on the considered examples. Moreover, this method provides a whole trajectory of couplings with increasing dependence. The same type of argument is applied to derivatives pricing: a European option on several underlyings is considered, and our dependence stress test is compared to the more classic stress test of covariance matrices, which typically assumes that the cross-covariance matrix is filled with a single parameter ρ and let ρ tend to ± 1. This last method has a major disadvantage when the marginals are fixed: the covariance matrix has two fixed diagonal blocks (the covariance matrices of the marginals), and the parameter ρ is constrained to belong to an interval to ensure the nonnegativity of the covariance matrix. It results that our method has a larger impact on the prices and avoids the problem of maintaining the stressed covariance matrix in the set of symmetric nonnegative matrices.

Chapter 3 The third chapter tackles the issue of describing the dependence between stochastic processes with copulas, and shows how copulas can be used in a genuine dynamic framework. The point is to be able to describe the cross-sectional dependence between two Markovian diffusions X_i, $i = 1, 2$, whose dynamics are $dX_t^i = \mu_i(t, X_t^i)dt + \sigma_i(t, X_t^i)dW_t^i$. The coupling between these diffusions is materialized by a coupling correlation $\rho(t, X_t^1, X_t^2)$ between the Brownian motions W^i, these latter being defined in such a way that $d\langle X_t^1, X_t^2\rangle = \rho(t, X_t^1, X_t^2)dt$, i.e. $\rho(t, X_t^1, X_t^2)$ is the instantaneous correlation between the Brownian motions. Such models are reminiscent of Dupire's local volatility model [2], and can be used in the same manner, that is to calibrate the correlation function in order to match today's

prices of options; this is the approach of the local correlation models introduced by Langnau [6] and Reghai [10]. The problem here is a bit different as we want to adjust the correlation function in such a way that the copula C_t between X_t^1 and X_t^2, which fully describes the cross-sectional dependence at time t, is controlled. In the spirit of the copula approach to dependence, we describe the dynamics of the copula C_t by first scaling the marginal diffusions X_t^i by their cumulative distribution functions F_t^i: a pair (U_t^1, U_t^2) of stochastic processes with stationary uniform distribution obtains, whose bivariate cumulative distribution function is, by definition, the copula C_t. The dynamics of the copula C_t is then derived by establishing the Kolmogorov forward equation of the process (U_t^1, U_t^2). This equation writes

$$\begin{aligned}\partial_t C_t(u,v) = \frac{1}{2}\big(&vol_1(t,u)\partial^2_{uu}C_t(u,v) + vol_2(t,v)^2\partial^2_{vv}C_t(u,v)\big)\\ &+ \tilde{\rho}(t,u,v)vol_1(t,u)vol_2(t,v)\partial^2_{uv}C_t(u,v)\end{aligned}$$

where $vol_i(t,\cdot) = (f_t^i.\sigma_i(t,\cdot))\circ(F_t^i)^{-1}(\cdot)$ is the volatility of the scaled marginal i and $\tilde{\rho}(t,u,v) = \rho(t,(F_t^1)^{-1}(u),(F_t^2)^{-1}(v))$ is the scaled correlation function. This equation describes the evolution of C_t, which depends on the marginal distributions and on the coupling correlation ρ. This equation offers a means to control the copula of the bivariate diffusion. For a fixed family $\{C_t\}_{t\geq 0}$ with smooth and positive densities, define:

$$\tilde{\rho}(t,u,v) = \frac{\partial_t C_t(u,v) - \frac{1}{2}\big(vol_1(t,u)\partial^2_{uu}C_t(u,v) + vol_2(t,v)^2\partial^2_{vv}C_t(u,v)\big)}{vol_1(t,u)vol_2(t,v)\partial^2_{uv}C_t(u,v)} \tag{8}$$

and $\rho(t,x,y) = \tilde{\rho}(t,F_t^1(x),F_t^2(v))$. If ρ is indeed a correlation function, i.e. if $|\tilde{\rho}(t,u,v)| \leq 1$ for all $(u,v) \in [0,1]^2$, a sensible expectation is that the copula family of the bivariate diffusion with coupling correlation $\rho(t,x,y)$ is indeed C_t. The first part of this chapter consists in establishing the copula PDE as well as this result of 'coherence'.

In a second part, the emphasis is put on the simplest coupling case, which corresponds to Brownian motions coupling. The coupling stochastic differential equation for Brownian motions writes

$$dB_t^2 = \rho(t,B_t^1,B_t^2)dB_t^1 + \sqrt{1-\rho(t,B_t^1,B_t^2)^2}dZ_t \tag{9}$$

where $(B^1,Z)_t$ is a standard bivariate Brownian motion. The problem is to determine whether a given copula C is attainable by coupled Brownian motions, in other words, whether there exists a correlation function ρ such that the equation (9) makes sense and the resulting bivariate process (B_t^1,B_t^2) has a copula family C_t satisfying $C_T = C$ at some time $T > 0$. Furthermore, we are primarily interested in the copulas that are stationary, that is the copulas C such that there exists coupled Brownian motions (B_t^1,B_t^2) whose copula family C_t becomes constant and equal to C after some time $T > 0$. It turns out that the property of self-similarity and invariance under time-inversion of the Brownian motion considerably simplifies the analysis of

attainability, as a copula that is attainable at some time $T > 0$ is attainable at any time $t > 0$ (and the same for stationary copulas). In order to show that the set of stationary copulas does not reduce to the Gaussian copula, a detailed example is provided which shows that non trivial members of the Farlie-Gumbel-Morgenstern copula family are indeed stationary copulas. This analysis is followed by a short zoology of copulas, which aims at listing some classic copulas that are admissible or not as stationary copulas. The coupling correlation function of these copulas equals

$$\rho(u,v) = -\frac{1}{2}\frac{e^{\frac{\Phi^{-1}(v)^2-\Phi^{-1}(u)^2}{2}}\partial^2_{uu}C + e^{\frac{\Phi^{-1}(u)^2-\Phi^{-1}(v)^2}{2}}\partial^2_{vv}C}{\partial^2_{uv}C} \tag{10}$$

after some time $T > 0$, according to (8). Empirically, and for the copula we tested, the copulas seem to divide in two categories, the ones with $\sup_{(u,v)\in[0,1]^2} |\rho(u,v)| \leq 1$ and the ones such that $\rho(u,v)$ explodes when (u,v) is close to the boundary of the unit square. These latter copulas are not stationary copulas, and numerical evidence suggests that classic copulas such as the Student copula, or many archimedean copulas have this behavior. Furthermore it is worth noticing that all copulas with bounded correlation are copulas without tail dependence, like the Gaussian copula, while the copulas for which $\rho(u,v)$ explodes exhibit tail dependence. This might prevent them from coupling Brownian motions.

The final part of this chapter is devoted to a financial application of the previous coupling method. The impact of introducing non Gaussian dependence is assessed on a dynamic strategy, namely a CPPI Long-Short strategy. This strategy involves two assets, the core and the satellite, and aims at guaranteeing a wealth that is proportional to the buy-and-hold strategy in the core, while benefiting from a possible rise of the satellite. This is achieved by alternatively shorting one asset and being long the other according to the value of the strategy relative to the value of the guarantee. The diffusion model for the assets is a coupled Black-Scholes, and we focus on the gap risk, materialized by the probability that the Long-Short CPPI falls below the buy-and-hold strategy. The impact of copulas is monitored: although relatively low (compared to the sensitivity of the gap risk with respect to the volatility for instance), it is real and shows that some models of dependence are more conservative than others for the strategy under consideration.

Bibliography

[1] W. Darsow, B. Nguyen, and E. Olsen. Copulas and Markov processes. *Ill. J. Math.*, 36(4):600–642, 1992.

[2] B. Dupire. Pricing with a Smile. *Risk*, 1994.

[3] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence properties in risk management: Properties and pitfalls. In M. Dempster, editor, *Risk Management: Value at Risk and Beyond*. Cambridge University Press, 2002.

[4] C. Genest, J. Q. Molina, and J. R. Lallena. De l'impossibilité de construire des lois à marges dimensionnelles données à partir de copules. *C R Acad Sci Paris Sér I Math*, 320:723–726, 1995.

[5] R. Ibragimov. Copula-based characterization for higher-order Markov processes. Technical report, Department of Economics, Harvard, 2005.

[6] A. Langnau. A dynamic model for correlation. *Risk*, April 2010.

[7] D. X. Li. On default correlation: A copula function approach. *Journal of Fixed Income*, 9(4):43–54, 2000.

[8] L.Rüschendorf. Law invariant convex risk measures for portfolio vectors. *Statistics & Decisions*, 24:97–108, 2006.

[9] A. J. Patton. Copula-based models for financial time series. In T. Anderson, R. Davis, J.-P. Kreiss, and T. Mikosh, editors, *Handbook of Financial Time Series*. Springer-Verlag Berlin Heidelberg, 2009.

[10] A. Reghai. Breaking correlation breaks. *Risk*, October 2010.

[11] P. Schönbucher. *Credit derivatives pricing models*. Wiley Finance, 2003.

[12] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics. American Mathematical Society, 2003.

Chapter 1

Numerical approximation of optimal transport maps

1.1 Introduction

The problem of interest is to compute numerically a solution to the Monge-Kantorovich problem in L^2. It consists in finding a law of probability π of a pair of random variables (X, Y) with marginal distributions μ and ν over $\mathbf{R}^n$ that solves

$$\mathbf{E}_\pi(X \cdot Y) = \max_{\tilde{\pi} \in \Pi(\mu,\nu)} \mathbf{E}_{\tilde{\pi}}(X \cdot Y) \tag{1.1}$$

where $\Pi(\mu, \nu)$ is the space of all such joint laws, $\mathbf{E}_\pi$ denotes the expected value with respect to the law π and $\cdot$ is the Euclidean scalar product. This problem has received a considerable attention. Originally formulated by Gaspard Monge in a stronger form in 1781, and under the above form by Leonid Kantorovich in the forties, it has found many applications in many fields, both theoretical and practical. Excellent references are the two volumes by Rachev and Rüschendorf [18] as well as the books by Cédric Villani [22] and [24] that show the phenomenal scope of the optimal transport theory.
The *primal problem* (1.1) is equivalent to the *dual problem*

$$\inf_{\varphi \in L^1(d\mu)} \int \varphi d\mu + \int \varphi^* d\nu \tag{1.2}$$

where φ^* stands for the Legendre transform of φ,

$$\varphi^*(y) = \sup_{x \in \mathbf{R}^n} [x \cdot y - \varphi(x)] \tag{1.3}$$

Under light assumptions on the marginals μ and ν, Brenier's theorem [7] states that there is a unique solution to the primal problem, and this solution is the law of a couple $(X, \nabla\varphi(X))$, with second marginal is a deterministic function of the first one. Moreover φ, called a *Kantorovitch potential* , is a solution to the dual problem (1.2) that is convex, and $\nabla\varphi$ is the unique μ- a.e gradient of a convex function such that $\nabla\varphi\#\mu = \nu$.
The optimal transport map $\nabla\varphi$ is a rather complicated object. It is a solution of the highly non-linear Monge-Ampère equation

$$\det D^2\varphi(x) f_\nu(\nabla\varphi(x)) = f_\mu(x)$$

when μ and ν have densities f_μ and f_ν with respect to the Lebesgue measure.

The numerical side of optimal transport has received less attention compared to the numerous theoretical developments. Notable exceptions include Brenier and Benamou [3] who derive a saddlepoint formulation of the transport problem and make use of augmented Lagrangian techniques to propose estimates for the optimal transport map; they present results when the initial measure is uniform on the torus $\mathbf{R^n}/\mathbf{Z}^n$. Loeper and Rapetti [15] solve the Monge-Ampère equation (with constant right term) $D^2\psi = \rho$, where ρ is a smooth density by using a linearization of the equation combined with a Newton's algorithm. Results are provided again in the case where ρ is the uniform measure on the torus. Angenent, Haker and Tannenbaum [1] and Dominitz *and al.* [9] use the equivalent problem of polar factorization to design a gradient-descent algorithm.

Another strand of literature that is not directly connected to the determination of the optimal transport maps deals with optimal transport when the marginals are discrete. When they have the same number of equally-weighted atoms, this is the *assignment problem* . This is the classical matching problem of assigning N people to N objects while maximizing a matching function c (the scalar product in the case of L^2 optimal transport). The problem reduces to finding a permutation of N elements that solves

$$\max_\sigma \sum_i c_{i\sigma(i)}$$

where the maximization is performed over the set of permutations of $\{1, \dots, N\}$.
This problem has been extensively studied in combinatorial optimization. An important contribution is the auction algorithm, proposed by Bertsekas [5], that is to our knowledge the most efficient algorithm for solving the assignment problem.

In a first part, the transport problem with discrete target measure is investigated. A simple but enlightening economical interpretation is given. The approach proposed is related to power diagrams and least-square clustering: Aurenhammer et al. [2] detail a gradient-descent

algorithm for solving the least-square assignment problem. We then provide (under appropriate assumptions on the marginal measures) theoretical pointwise convergence results stating that the Kantorovich potential in the case of discrete target measure ν_N converges uniformly to the Kantorovich potential of the dual problem as soon as ν_N converges in distribution to ν. Eventually, a gradient-descent algorithm is proposed as well as empirical experiments. We compare this type of gradient-descent algorithm coupled with a quick warm-point algorithm to linear programming algorithms as well as to the popular auction algorithm, and show that it can be an efficient means to solve the transport problem.

1.2 Related literature and contribution

The algorithm presented in the first part of this chapter and compared to other algorithms in the last part is close to a solution to the least-square assignment problem proposed in Aurenhammer, Hoffmann and Aranov [2]. This problem can be formulated in the following way: find a polyhedral partition of the space $\mathbf{R}^n$ by power diagrams with given volume. A power diagram is a partition of the space into 'distorted Voronoï cells' , that can be written as

$$P_i(v) = \{x|\ |x - y_i|^2 - w_i \leq |x - y_j|^2 - w_j,\ j = 1, \ldots, M\}$$

where the M points $y_i \in \mathbf{R}^n$ are the sites of the diagram and $w_i \in \mathbf{R}$ are the weights. These are the analogue of the places and the prices of the section 1.3.3. The problem is to find, for a given vector of 'capacities' c (that corresponds to the offer of the sellers, q, in 1.3.3), and a probability distribution μ over $[0,1]^n$, a vector of weights w such that $c_i = \mu(P_i(v))$ for all i. This is almost exactly the problem of finding the optimal transport map between μ and $\sum_{i=1}^M c_i \delta_{y_i}$ as explained in the next section 1.3; this was pointed out by Rüschendorf and Uckelmann [20]. Aurenhammer *et al.* show that the optimal weights are the maximum of a concave function, just as the optimal prices of section 1.3.3 are the minimum of a convex function, and propose to compute these weights by a gradient method, which is the analogue of the method we use. Gangbo and McCann [11] conjectured that such an algorithm should yield a solution. The results given in the section 1.3 are a direct extension of those presented in Ekeland, Galichon and Henry [10], section 3, who cover the case of a discrete initial distribution and provide an economic interpretation close to the one exposed in 1.3.3.

The second part of the chapter consists in proving that the optimal transport map between two continuous measures can be approximated by solving the transport problem when the target measure ν is approached by a sequence ν_N converging in distribution to ν. This result has been proved in quite a general setting by Villani [23], and we provide here a detailed proof that fits into the setting of this chapter (optimal transport over the Euclidean space $\mathbf{R}^n$).

The third part of this chapter compares several algorithms, and especially the classic auction

algorithm of D. Bertsekas (see e.g. [5]) taken as the benchmark. The Iterative Proportional Fitting Procedure (IPFP) is proposed as a means to provide a warm point for the gradient descent algorithm. A similar entropic relaxation problem is studied, with no apparent connection made with the transport problem, by Yuille and Kosowsky [13], who observe (as we do) the lack of performance of this algorithm when the discretization of the measures involves a large number of atoms. Finally, let us insist on the fact that we chose to solve the transport problem with measures on the Euclidean space, while some authors (e.g. Loeper and Rapetti [15]) solve it on the torus; it seems that, in this case, the convergence of the various algorithms is way faster than it would have been in the Euclidean setting.

1.3 Optimal transport with a discrete target measure

In what follows, μ and ν are always supposed to have compact support and do not give mass to small sets [1], so that Brenier theorem mentioned above applies and the optimal transport map is well defined. The support of μ is supposed connected so that Kantorovitch potential is defined up to a constant. The case of transport problem between μ and a discrete measure of probability with a finite number of atoms is well-known and admits explicit solutions.

1.3.1 Form of the solution

Let ν_N be a discrete probability over $\mathbf{R}^n$, with N atoms, $\nu_N = \sum_{i=1}^N q_i \delta_{y_i}$. A mapping ψ pushing μ forward to ν_N satisfies

$$\psi(X) \sim \sum_{i=1}^N q_i \delta_{y_i}, \text{ when } X \sim \mu$$

where $\sim$ means equality in distribution.
This implies that on suppμ, ψ takes values in the finite set $\{y_i,\}_{1\leq i\leq N}$. Actually, according to Brenier Theorem, we know that there exists a μ-a.e unique gradient of a convex function which solves the problem: the previous remark indicates that this convex function should be looked for under the form of a piecewise affine function on suppμ. More precisely:

Proposition 1 *An optimal convex function φ satisfying $\nabla\varphi\#\mu = \nu_N$ is piecewise affine, i.e. the Brenier Map $\nabla\varphi$ is a piecewise constant function. More explicitly, a solution φ is*

$$\varphi(x) = \max_{i=1,\dots,N} [x \cdot y_i - v_i] \tag{1.4}$$

Let $v = (v_1, \dots, v_N)'$ and $V_i(v) = \{x \in supp\mu | i \in argmax_k [x \cdot y_k - v_k]\}$. The v_i are such that

$$\mu(V_i(v)) = q_i, \text{ for all } i$$

[1] A small set is a measurable subset of $\mathbf{R}^n$ with Hausdorff dimension at most $n-1$.

Proof: Let φ_v be as in (1.4), its gradient is given by

$$\nabla_x \varphi_v(x) = \sum_{i=1}^{N} \mathbb{1}_{V_i}(x) y_i, \; a.e.$$

Thus

$$\nabla_x \varphi_v \# \mu = \sum_{i=1}^{N} \mu(V_i) \delta_{y_i}$$

thanks to the envelope theorem (here and thereafter the dependence on v in V_i is often dropped). A necessary and sufficient condition on $(v_i)_{1 \leq i \leq N}$ for $\nabla \varphi_v$ to solve the problem is:

$$\mu(V_i) = q_i \text{ for all } i. \tag{1.5}$$

□

However, it is not clear at this point whether such a set of v_i actually exits. Theorem 1.4 and example 1.6 in [11] state that this is indeed the case. Alternatively, the proposition 2 below proves also the existence of such a solution.

The actual computation of the optimal transport map reduces to finding such a v; the dual transport problem provides a means to do this.

1.3.2 Dual problem and first order condition

The Kantorovich potential is a solution of the dual problem

$$\inf_{\varphi \in L^1(d\mu)} \int \varphi(x) d\mu + \int \varphi^*(y) d\nu_N$$

Thus, φ_v is optimal iff it is a solution of the following minimization problem :

$$\inf_{v \in \mathbf{R}^N} \int \varphi_v(x) d\mu + \int \varphi_v^*(y) d\nu_N(y)$$

A straightforward computation yields $\varphi_v^*(y_j) = v_j$. Thus the dual problem writes

$$\inf_{v \in \mathbf{R}^N} \left[\int \varphi_v(x) d\mu + q \cdot v \right] \tag{1.6}$$

Formally, $\nabla_v \mathbf{E}(\varphi_v(X)) = \mathbf{E}(\nabla_v \varphi_v(X)) = -(\mu(V_1) \dots \mu(V_N))'$, and the first order condition reads $q_j = \mu(V_j), \; j = 1, \dots, N$ which is precisely the condition (1.5) for φ_v to be a Kantorovich potential.

A remark that is the cornerstone of the algorithm proposed below is the convexity of the objective function $\mathcal{F}(v) := \mathbf{E}(\varphi_v(X)) + q \cdot v$.

Proposition 2 *$\mathcal{F}$ is a convex function that is bounded below and thus admits a global minimizer in $\mathbf{R}^N$. Moreover, $\mathcal{F}$ is continuously differentiable on $\mathbf{R}^N$ and any global minimizer v satisfies the first order condition*

$$q_i = \mu(V_i(v)), \text{ for all } i$$

A proof can be found in appendix 1.7.1.

1.3.3 An economic interpretation

A simple economic interpretation gives some insight into the transport problem. Consider a spatial distribution μ of buyers interested in a single type of good, supplied by sellers located at positions y_i. q_i models the offer proposed by the i-th seller. Each customer faces a trade-off between a distance cost and the prices proposed by the sellers. The economic surplus of assigning the buyer located at x to the i-th seller (the one located at y_i) is set as $x \cdot y_i$. The primal problem $\max_{X\sim\mu, Y\sim\nu_N} \mathbf{E}(X \cdot Y)$ consists in the maximization of the total economic surplus, and is the problem the social planner wants to solve: controlling the coupling between the distribution of the customers and the distribution of the sellers, so as to maximize the total surplus. Welfare theorems suggest that this problem should be related to some price equilibrium (Walrasian equilibrium) that arises from the price competition between the sellers. Indeed, the first order condition of the dual problem states that $\mu(V_i) = q_i$ for all i. Recall that:

$$V_i = \{x | x \cdot y_i - v_i \geq x \cdot y_j - v_j, \forall j\}$$

In other words, the set V_i is the set of customers whose net surplus is maximum when they buy from seller i: this is in some way the basin of attraction of the seller i. Hence, $\mu(V_i)$ is the proportion of customers buying from the i-th seller. The dual problem amounts to adjusting the prices v_i so that the proportion $\mu(V_i)$ equals q_i, the offer supplied by the i-th seller, that is *equalizing supply and demand*. Thus the dual problem consists in adjusting the prices so that the equilibrium between supply and demand is attained. Therefore, market clearing is equivalent to maximizing total surplus. Finally, remark that the gradient-descent algorithm proposed below implements the Walrasian auction process: sellers that sell their whole production raise their prices, reducing the size of their basin of attraction, while those who fail at selling their entire offer lower their prices, increasing their basin of attraction. The process repeats until the market clears (possibly within an infinite amount of time).

1.4 Approximating the continuous case

The solution to the problem in the continuous case draws upon the previous results by discretization of the target measure: let ν_N be any sequence of discrete probability measures

converging in law to ν. For each N, there exists a solution $\nabla\varphi_N$ to the optimal transport problem between μ and ν_N. The problem of interest is to give sufficient conditions that ensure the convergence of the sequence $\nabla\varphi_N$ to $\nabla\varphi$ and that justify the two steps approach 'discretize the target measure and take limits'.

Convergence in measure in L^2 A first result concerning the convergence of the optimal transport map is given in [24] Corollary 5.23:

$$\text{For all } \varepsilon > 0,\ \mu\{x : |\nabla\varphi_N(x) - \nabla\varphi(x)| > \varepsilon\} \to 0$$

i.e. the convergence *in measure* of $\nabla\varphi_N$ to $\nabla\varphi$ with respect to μ. This result holds under fairly general assumptions on μ and ν (compactness is not necessary). However, this result does not tell anything about the possible pointwise convergence of φ_N or its gradient, but the existence of a subsequence $\nabla\varphi^{N_j}(X)$ that converges almost surely to $\nabla\varphi(X)$ as $j \to +\infty$. A similar result holds for the optimal transport plans π_n when both initial and target measures are approximated by sequences of discrete measures as proved in the first volume of [18].

Pointwise convergence A stronger form of convergence can be proved under additional hypothesis. Pointwise convergence results are less known, see for instance an article by Villani [23]. The proof proposed below is adapted from this article to our simpler setting. This proof follows the strategy of finding a uniformly converging subsequence of φ_N, and proving that any such subsequence converges to the optimal transport map. The existence of a converging subsequence relies is ensured by Ascoli's theorem. In order to invoke this theorem we have to make sure that the sequence $\{\varphi_k\}$ in $\mathcal{C}(\operatorname{supp}\mu)$ is uniformly bounded. However in our setting where μ and ν have a compact support, this is an easy lemma, proved in appendix 1.7.1.

Lemma 1 *The sequence of vectors $\{v_N\}$ are uniformly bounded, as well as the sequence $\{\varphi_N\}$.*

As a consequence of Ascoli theorem (see appendix 2):

Lemma 2 *There exists a subsequence $\{\varphi_{k'}\}$ which converges in $\mathcal{C}(\operatorname{supp}\mu)$.*

The convergence is proved under the following hypothesis **(H)**

1. μ and ν have compact and convex support.
2. They do not give mass to small sets.
3. μ is absolutely continuous with respect to the Lebesgue measure λ and $\frac{d\mu}{d\lambda} > 0$ almost everywhere on the support of μ.

As we will see, the limits of the converging subsequences are defined up to constant. To fix things, every φ_k is supposed w.l.o.g to be zero at a fixed point x_0 in the support of μ.

Theorem 1 *Under* ***(H)****,* φ_k *converges uniformly to a convex function whose gradient (defined almost everywhere) is the optimal transport map.*

In the following proof, the value of the functions outside the supports of μ or ν is not important. However, they must be defined everywhere (and possibly convex), and this is done by giving them the value $+\infty$ outside these supports.

Proof: The proof consists in proving first that every converging subsequence has a limit whose gradient is the optimal transport map. Then, we show that these functions are equal up to an additive constant, that is zero up to a normalization of φ_N. This concludes the proof as the existence of a subsequence and the uniqueness of the limits of the converging subsequence entails the convergence of the sequence $\{\varphi_N\}$.

Let $\{\varphi_{k'}\}$ a subsequence of $\{\varphi_k\}$, and φ its limit in $\mathcal{C}(\operatorname{supp}\mu)$. Let us prove that the gradient of φ is a solution of the optimal transport, i.e. $\nabla\varphi\#\mu = \nu$.

The restriction of φ_k^* to $\operatorname{supp}\nu$ enjoys the same properties as φ_k: they are uniformly bounded and equi-Lipschitz. Thus Ascoli theorem implies that there exists a subsequence still noted k' such that both $\varphi_{k'}$ tends to φ and $\varphi_{k'}^*$ tends to a function (continuous and convex) ψ, respectively uniformly over $\operatorname{supp}\mu$ and $\operatorname{supp}\nu$.

The quantity

$$\int |x|^2/2 - \varphi_{k'}(x)d\mu(x) + \int |y|^2/2 - \varphi_{k'}^*(y)d\nu(y)$$

tends to

$$\int\limits_{\operatorname{supp}\mu} |x|^2/2 - \varphi(x)d\mu(x) + \int\limits_{\operatorname{supp}\nu} |y|^2/2 - \psi(x)d\nu(y)$$

by dominated convergence. However, the upper expression is $\mathcal{W}_2^2(\mu,\nu_{k'})$ the quadratic Wasserstein distance between μ and $\nu_{k'}$. According to the continuity of the Wasserstein distance (Theorem 7.12 in [22]), $\mathcal{W}_2^2(\mu,\nu_{k'}) \to \mathcal{W}_2^2(\mu,\nu)$ and thus

$$\int\limits_{\operatorname{supp}\mu} |x|^2/2 - \varphi(x)d\mu(x) + \int\limits_{\operatorname{supp}\nu} |y|^2/2 - \psi(x)d\nu(y) = \mathcal{W}_2^2(\mu,\nu) \tag{1.7}$$

which means that the couple $(|x|^2/2-\varphi, |y|^2/2-\psi)$ is optimal in the dual Monge-Kantorovich problem. However, the solution of the dual problem is not necessarily unique, and we have to prove that $\psi = \varphi^*$ in order to conclude.

It is sufficient to prove that $\psi = \varphi^*$ *on* $\operatorname{supp}\nu$. As $\varphi_{k'}(x) + \varphi_{k'}^*(y) \geq x\cdot y$, taking limits yields

$$\varphi(x) + \psi(y) \geq x\cdot y,\ \forall x \in \operatorname{supp}\mu, \forall y \in \operatorname{supp}\nu \tag{1.8}$$

As φ is infinite outside $\operatorname{supp}\mu$, $\varphi^*(y) = \sup_{x\in\operatorname{supp}\mu}[x\cdot y - \varphi(x)]$ and (1.8) implies that

$$\psi \geq \varphi^* \text{ on } \operatorname{supp}\nu \tag{1.9}$$

If there exists $y \in Int(\operatorname{supp}\nu)$ such that $\psi(y) > \varphi^*(y)$, the continuity of ψ and φ^* on $\operatorname{supp}\nu$ implies that there exists a ball $B \subset \operatorname{supp}\nu$ over which $\psi - \varphi^* \geq \varepsilon > 0$. So,

$$\begin{aligned} \int \varphi d\mu + \int \psi d\nu &\geq \int \varphi d\mu + \int \varphi^* d\nu + \nu(B)\varepsilon \\ &> \int \varphi d\mu + \int \varphi^* d\nu \\ &\geq \inf_{\varphi \in L^1(d\mu)} \int \varphi d\mu + \int \varphi^* d\nu \end{aligned} \tag{1.10}$$

However,

$$\sup_{(\varphi,\psi)\in L^1(d\mu)\times L^1(d\nu)} \int \varphi d\mu + \int \psi d\nu = \underbrace{\int \frac{|x|}{2} d\mu + \int \frac{|y|}{2} d\nu}_{=M_2} - \inf_{\varphi \in L^1(d\mu)} \int \varphi d\mu + \int \varphi^* d\nu$$

But we know according to (1.7) that $M_2 - \int \varphi d\mu + \int \psi d\nu$ is a solution of the Monge-Kantorovich problem. Hence (1.10) is a contradiction and $\psi \leq \varphi^*$, on $Int(\operatorname{supp}\nu)$. This holds on $\overline{Int(\operatorname{supp}\nu)}$ by continuity. The convexity of $\operatorname{supp}\nu$ implies that $\overline{Int(\operatorname{supp}\nu)} = \operatorname{supp}\nu$ (according to the 'Accessibility Lemma', see [6]).
As a consequence, every converging subsequence has a limit whose gradient is the optimal transport map. This map is unique μ *a.e* and as μ is supposed absolutely continuous with respect to the Lebesgue measure with an almost everywhere positive density, it is unique almost everywhere on $\operatorname{supp}\mu$. This is sufficient to ensure that the limits of the converging subsequence are defined up to a constant. □

So far, we have seen convergence results for the Kantorovitch potential. This implies convergence of the optimal transport maps $\nabla\varphi_N$ in the case where the limit Kantorovitch potential φ is regular enough, as pointed out by Villani [23]. Without entering into the details of the Caffarelli's regularity theory (see Th. 4.14 [22] or Th. 12.50 [24]), if both supports are convex and regular enough and that μ and ν are absolutely continuous with respect to the Lebesgue measure with smooth and locally bounded from below densities, then the Kantorovitch potential is regular. If it is $\mathcal{C}^1$, we have

Lemma 3 (recalled in [23]) *Let φ be a convex function in $\mathcal{C}^1(\mathbf{R}^n, \mathbf{R})$ and φ_k a sequence of convex functions converging pointwise to φ on $\mathbf{R}^n$. Then if $C \subset \mathbf{R}^n$ is an open convex set on which φ is finite, then $\partial\varphi_k$ converge to $\partial\varphi$ locally uniformly on the compact subsets of C in the sense that*

$$\forall S \text{ compact subset of } C, ||d(\partial\varphi_k(\cdot), \nabla\varphi(\cdot))||_{+\infty, S} \underset{k\to+\infty}{\longrightarrow} 0$$

where d is the Euclidean distance between a set and a point and $\partial\varphi(x)$ denotes the subdifferential of φ at point x.

In other words, not only does φ_N converge locally uniformly to φ, but the subdifferential $\partial\varphi_N$ converges uniformly on the compacts of $\operatorname{supp}\mu$ to the optimal transport map.

1.5 Numerical Experiments

We first describe how is chosen the discretization of the target measures. A description of the algorithms is then provided and examples allows for a comparison of their respective performances.

1.5.1 Discretization of ν

We have not imposed any restriction so far as to the form of the approximating sequence ν_k except that it must converge in distribution to ν, and the above results hold for any such sequence.

The purpose of *optimal quadratic quantization* consists in finding a discrete probability ν_N with N atoms that is closer among such measures to a given probability ν, with respect to the quadratic Wasserstein distance . In other words, if $\mathcal{P}_N$ is the set of discrete measures with N atoms, then an optimal quadratic N-quantizer ν_N is a measure that solves

$$\min_{Q\in\mathcal{P}_N} \mathcal{W}_2^2(\nu, Q) = \min_{Q\in\mathcal{P}_N} \inf_{\pi\in\Pi(Q,\nu)} \mathbf{E}_\pi(|X-Y|^2)$$

It can be shown that such a measure takes the form $\nu_N(x) = \sum_{i=1}^N \nu(C_i(x))\delta_{x_i}$ where $x = (x_1,\dots,x_N) \in (\mathbf{R}^n)^N$ is called an optimal quadratic quantizer , and

$$C(x_i) = \{x; |x-x_i| \le |x-x_j|,\ j=1,\dots,N\}$$

are the Voronoï cells associated to x.
The discretization ν_N converges indeed to ν as implied by the following theorem (whose proof as well as an extensive exposition of the theory can be found in [21]):

Theorem 2 (Graf-Luschgy) *If ν has a finite moment of order strictly above 2, then*

$$\mathcal{W}^2(\nu, \nu_N(x))_{N\to\infty} \sim \frac{C_{n,\nu}}{N^{2/n}}$$

for some constant $C_{n,\nu}$ which depends on n and ν.

Thus to use such an approximation we need two things: an optimal N-quantizer $(x_1,\dots,x_N) \in \mathbf{R}^n$ and the weights of the Voronoï cells $\nu(C(x_i))$. The advantage of this approximation is can be computed once and for all for a given ν.

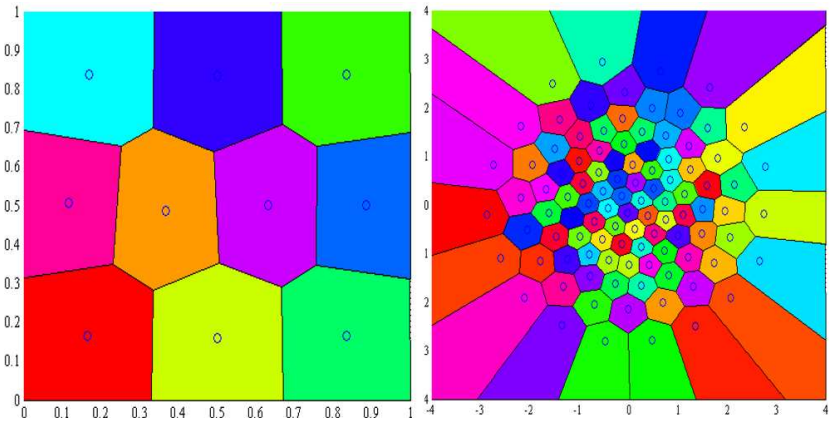

Figure 1.1: Left: optimal quantization of the uniform law on the square $[0,1]^2$. Right: optimal quantization of the bivariate normal law. This latter (as well as the optimal quantization grid of the normal law used in what follows) is taken from G.Pagés and J. Printems website [17].

1.5.2 The algorithms

We make the distinction between two classes of algorithms solving the transport problem. On the one hand, the algorithms that use the discretization of the original measure μ to reduce the problem to a linear programming problem or that use a very specific form of the discretization (for instance with equally weighted atoms) such as the auction algorithm. On the other hand, the minimization problem (3.16) does not impose a discretization of the initial measure. For instance, when μ is an uniform measure on a bounded polytope, it is not necessary to perform a discretization of μ to obtain a numerical solution. Furthermore, even if discretization of the target measure is chosen, it can take various forms: equally weighted atoms or not, same number of atoms as in the discretization of ν or not.

Continuous to discrete case

The dual problem formulation (3.16) is a convex optimization problem. As such it is natural to try a Newton-type algorithm that updates the prices iteratively according to:

$$v_{i+1} = v_i - t_i H_i^{-1} \nabla_v \mathcal{F}(v_i) \tag{1.11}$$

where H_i is an approximation of the Hessian of $\mathcal{F}$ at v_i. The gradient of $\mathcal{F}$ is given by

$$\nabla_v \mathcal{F}(v) = q - \begin{pmatrix} \mu(V_1(v)) \\ \vdots \\ \mu(V_N(v)) \end{pmatrix} \tag{1.12}$$

Descent algorithms The BFGS (Broyd, Fletcher, Goldfarb and Shanno) algorithm provided by the Optimization Toolbox in MATLAB® R14 (2005) is used. It falls into the broader category of quasi-Newton methods that aim at using the curvature of the objective function. It provides is a sequence B_k of semidefinite positive matrices that hopefully provides a good approximation of the inverse of the Hessian H_k^{-1} after a few steps and converges to the true Hessian inverse at the optimum, see [16]. B_0 can be set arbitrarily (it is the identity matrix in what follows) and a line-search is performed to determine the best step α_k.

Evaluation of the mass $\mu(V_i)$ This is the crucial point, and the most time consuming. When μ is a uniform law, it amounts to compute the volume of V_i; if suppμ is a polytope, so is V_i and triangulation techniques can be used to compute this volume accurately. However when the law μ differs from the uniform law, a discretization of μ is used.

Initial guess Finally, a crucial parameter in descent algorithm is the starting point. An heuristics consists in remarking that when linearizing the optimal transport (in the case where μ and ν are close), the order zero term is $|x|^2/2$. As $\varphi_v^*(y_j) = v_j$, it is sensible to start from $v_j = |y_j|^2/2 = \left(\frac{|\cdot|^2}{2}\right)^*(y_j)$, although it can be suboptimal in many situations. This choice amounts to taking the initial $V_i(v_0)$ as the Voronoi cells associated to $(y_i)_{1\leq i\leq N}$.

Discrete μ

In this section, μ is the discrete distribution $\sum_i p_i\delta_{x_i}$. We tested three quite distinct solutions.

Linear programming methods The primal problem writes:

$$\min_{\forall i,j\ \varphi_i+\psi_j\geq x_i\cdot y_j} \sum_{i=1}^N p_i\varphi_i + \sum_{j=1}^N q_j\psi_j$$

This is a *linear programming* problem: the objective is linear, as well as the inequality constraints. Two standard algorithms are at our disposal, the common simplex algorithm, and primal-dual methods, also known as interior point methods. This last algorithm shall have our preference, as it is known to perform better than the simplex on large-scale problems. It is closely related to unconstrained linear programming with log-barrier penalization (see [16]). The set of constraints in this particular problem is huge: if there are N atoms, the number of constraints is N^2. As we will see, this feature ruins the performance of both algorithms when the number of points increases.

The auction algorithm This popular algorithm has been proposed by Bertsekas , see for instance the survey paper [4]. It solves the dual problem :

$$\min_{v\in\mathbf{R}^N}\left\{\sum_{i=1}^{N}\max_{j=1,\dots,N}[x_i\cdot y_j-v_j]+\sum_{j=1}^{N}v_j\right\} \tag{1.13}$$

This problem is the discrete version of the problem (3.16), in the case where μ and ν *are both equally distributed discrete measures*: this is the classic assignment problem . In a few words, the algorithm proceeds as follows. A set of N customers is to be assigned to N objects, in a one-to-one mapping. Each customer starts unassigned. Then each unassigned customer, say the i-th one, bids for the object j_i that maximizes its net surplus $s_i = \max_j(x_i\cdot y_j - v_j)$, his 'best object'. His bid is such that he becomes indifferent between this best object and the second best object j_i' (the object that maximizes $w_i = \max_{j\neq j_i}(x_i\cdot y_j - v_j)$): he is willing to increase the price of j_i by the bidding increment $s_i - w_i$. Once every unassigned bidder has made a bid, each object j is assigned to the bidder that has proposed the highest bidding increment for j. The person, if any, that was assigned to this object becomes unassigned and the price of object j rises by the highest bid. The process is repeated until everyone becomes assigned.

This would work fine if there were not prices war: when there is a group of objects that offers the same net surplus for some customer, the bidding increment may be zero. This can be the case for several customers that compete for equally desirable objects without raising their prices. Such cycles are broken by assuming that the bidding increments are replaced by $s_i - w_i + \epsilon$, $\epsilon > 0$, imposing a minimum increase of the price of a desired object by ϵ. This condition ensures that the algorithm ends and that the prices are within $N\epsilon$ to be optimal. The procedure becomes: for a given ϵ, unassign everyone and perform the auction procedure until everyone is assigned. Then lower ϵ and perform the auction process, starting with the prices that were found with the previous value of ϵ. Eventually, stop when ϵ falls below some threshold.

This algorithm is fast, in principle: the time complexity, for a given ϵ is $O(N^2 log(N.\max_{i,j}|x_i\cdot y_j|/\varepsilon))$, *when the surplus matrix* $(x_i\cdot y_j)_{i,j}$ *is integer*. In this case, the threshold for ϵ is fixed at $1/N$, because prices will then be within $N\epsilon < 1$ optimal, and then optimal because they are integer. This problem is overcome by a proper scaling: both x_i and y_j are multiplied by an integer and then floored. The solution v is then multiplied by the same integer thus giving an approximate solution of the initial problem.

An important remark is that this algorithm converges very well when the surplus matrix is sparse, but is less efficient with dense matrices, which is systematically the case in the following. The theoretical overall complexity (including the repeated ϵ-scaling) in the worst-case is $O(N^3\log(N.\max_{i,j}|x_i\cdot y_j|))$.

The algorithm tested here is the so-called forward auction algorithm (see [4]), with the integer

scaling described above; ϵ is set at 1 at the beginning, the ϵ-scaling is chosen as the reduction by one fourth at each iteration, and the terminal condition is $\epsilon \leq 1/N$.
The idea behind the application of this algorithm is that although is does not exactly solve the problem we are interested in, its fastness makes it appealing combined to a Monte-Carlo approach. Namely, instead of using a 'clever' discretization of the measures such as determined by the quantization, they could be discretized as a sum of a large number of equally weighted atoms and the auction algorithm would perform the assignment on this set of atoms.

Simulated Annealing and the Iterative Proportional Fitting Procedure Simulated annealing consists in introducing an entropic perturbation in the primal problem, yielding a relaxed version of the problem :

$$\pi_T \in \underset{\tilde{\pi}\in\Pi(\mu,\nu)}{\operatorname{argmax}}\{\mathbf{E}_{\tilde{\pi}}(X\cdot Y) + TEnt(\tilde{\pi})\} \tag{1.14}$$

As μ and ν are both discrete here, the entropy is defined as $-\sum_{i,j}\pi_{ij}\log\pi_{ij}$. T is a temperature parameter, and as T goes to zero, the entropy penalization vanish and the probability π_T becomes an approximate for the optimal coupling that solves

$$\pi \in \underset{\tilde{\pi}\in\Pi(\mu,\nu)}{\operatorname{argmax}} \mathbf{E}_{\tilde{\pi}}(X\cdot Y)$$

Let $\pi_T^0 \propto e^{x\cdot y}$ be a probability density. Then is is straightforward to see that the relaxed problem (1.14) is equivalent to the problem

$$\min_{\pi\in\Pi(\mu,\nu)} D_{KL}(\pi|\pi_T^0), \quad \text{where } D_{KL}(\pi|\pi_T^0) = \mathbf{E}_\pi\Big(\log\Big(\frac{\pi(X,Y)}{\pi_T^0(X,Y)}\Big)\Big)$$

D_{KL} is called the Kullback-Leibler divergence . Thus the relaxed problem amounts to 'project' (in a broad sense, as D_{KL} is not a distance) π_T^0 onto the set $\Pi(\mu,\nu)$ with respect to the Kullback-Leibler divergence. It can be shown, and a detailed proof is given in [19], that the solution π_T has the following log-likelihood: $\log\pi_T(x,y) = \frac{x'y}{T} + a_T(x) + b_T(y)$, $a \in L^1(d\mu)$, $b \in L^1(d\nu)$.
The IPFP algorithm, also known as Deming and Stefan algorithm is the alternative projection algorithm applied to the previous projection problem. It consists in alternatively projecting π_T^0 on the set $\Pi(\mu)$ of probabilities on $\mathbf{R}^n \times \mathbf{R}^n$ whose first N-dimensional marginals are μ and $\Pi(\nu)$, the set of probabilities whose second N-dimensional marginal are ν. This provides a sequence π_n, such that $\pi_{2n} \in \Pi(\mu)$, $\pi_{2n+1} \in \Pi(\nu)$ and $\pi_n \to \pi_T \in \Pi(\mu)\cap\Pi(\nu) = \Pi(\mu,\nu)$, in total variation. The algorithm consists in alternatively modifying a_T and b_T so that $\pi_{2n} \in \Pi(\mu) \propto e^{\frac{x'y}{T}+a_T^n(x)+b_T^n(y)}$ and $\pi_{2n+1} \in \Pi(\nu) \propto e^{\frac{x'y}{T}+a_T^{n+1}(x)+b_T^n(y)}$

If we let $(p_1, \ldots, p_N)'$ be the atoms of μ and $(q_1, \ldots, q_M)'$ the atoms of ν, the iteration writes:

$$\left\{ \begin{array}{rcl} e^{b_T^{n+1}(y_j)} & = & \frac{q_j}{\sum_i \pi_T^0(x_i,y_j) e^{a_T^n(x_i)}} \\ e^{a_T^{n+1}(x_i)} & = & \frac{p_i}{\sum_j \pi_T^0(x_i,y_j) e^{b_T^{n+1}(y_j)}} \end{array} \right.$$

with $b_T^0 = a_T^0 \equiv 0$. Eventually, $T.b(y_j)$ approximates v_j as $T \to 0$. On this latter fact, we refer to an article by Kosowski and Yuille [12] that relates the IPFP to Sinkhorn theorem and to the Iterative Scaling Algorithm. We should emphasize the fact that, contrary to the auction algorithm, IPFP is meant to work with distributions μ and ν that do not necessarily have the same type of atoms (for instance, they are not necessarily equally weighted sums of Dirac distributions).

1.5.3 Examples

We have tested our algorithm on three examples in dimension 2. These examples are presented as a proof of concept, in so far as the theoretical optimal transport map is obvious in each case as they all involve a simple transformation of the marginals.

1. The transport is between two uniform distributions and consists in a translation: $\mu = \mathcal{U}_{[-6,-2]^2}$, $\nu = \mathcal{U}_{[-2,2]^2}$. The optimal transport map is $\nabla\varphi(x) = x + 4\mathbb{1}$ i.e $\varphi(x) = |x|^2/2 + 4\mathbb{1}'x + cst$. We use a cube quantizer:
$(-2 + 4\frac{(2i-1)}{2N}, -2 + 4\frac{(2j-1)}{2N})_{1 \leq i,j \leq N}$.

2. Departing from the theoretical framework described above, we investigate the case of a non compactly supported target measure. μ is the uniform law on the unit square while ν is the bivariate normal law. As the initial and target measures are the laws of independent variables, the optimal transport map is obtained a scaling of the marginals: $\nabla\varphi(x) = (\Phi^{-1}(x_1), \Phi^{-1}(x_2))$.

3. Dilatation of normal distribution : $\mu = \mathcal{N}(0, Id)$ and $\nu = \mathcal{N}(0, 2Id)$: $\nabla\varphi(x) = \sqrt{2}x$ i.e $\varphi(x) = \frac{1}{\sqrt{2}}|x|^2 + cst$.
Optimal quantization grids $(y_i)_{i=1,\ldots,N}$ for the normal law $\mathcal{N}(0, Id)$ are those provided by [17].

1.5.4 Results

The following numerical results were computed with Matlab 7.5 running on a Xeon CPU @ 3 Ghz. We detail the results obtained for each example, while putting the emphasis on the first one.

Uniform case

Numerical results for the first example are given in appendix 1.7.2. Before entering the details of each algorithm, we give here the conclusions as to the best performing methods in this case. Our experiments rule out the linear programming algorithms, that proved very costly. The IPFP algorithm is fast, even compared to the auction algorithm, but generates a numerical error that does not vanish on the boundary of the support of μ, and therefore is retained as a warm point provider for the descent algorithms. The auction algorithm is also fast (compared to the descent algorithm for a fixed number of points of the discretization of ν) but yields higher numerical error than the descent algorithm for comparable computing time. Eventually, the descent algorithm that uses the quantization of the initial measure suffers the drawback of providing a numerical error that is not a decreasing function of the number of points of the quantization grids. This feature is even more pronounced than in the case of the auction algorithm.
Our conclusion is that the BFGS algorithm coupled with a warm point provided by IPFP and an accurate computation of the volume of the cells V_i should be chosen, as it is faster than the auction algorithm.

Exact Descent In the uniform case we use the Multiparametric Toolbox [14] which allows for polytopes manipulation. In particular, $\mu(V_i) = \frac{vol(V_i)}{vol(\text{supp}\mu)}$ is computed with this toolbox. We call this method *exact descent* as it uses the BFGS algorithm to determine the descent direction and because the volume of the polytopes V_i is computed by triangulation techniques via the MPT Toolbox and does not rely on a discretization of the initial measure. Table 1.2 sums up the results in the uniform case. It contains the number of points used in the quantization of $\mathcal{U}_{[-2,2]^2}$, the computation time and the numerical error. This latter is defined as

$$\sup_{0\leq i,j\leq 100} \left|\varphi_{\text{theory}}\Big(\frac{i}{100},\frac{j}{100}\Big) - \varphi_{\text{num}}\Big(\frac{i}{100},\frac{j}{100}\Big)\right|$$

Auction and IPFP Algorithm We use a regular grid[2] to discretize the uniform law both in the auction and IPFP algorithm. Table 1.4 shows it is extremely fast compared to the gradient algorithm for a fixed number of points. However, the exact descent is more efficient, as for comparable computation times, it yields a better numerical error. The IPFP is even faster than the auction algorithm (see table 1.5). However, when the number of points increases, the error does not necessarily decrease: it remains high on the boundary of the domain, and does not decrease even when the size of both discretizations of μ and ν increases. This is likely due to the fact that this algorithm fails at converging when the temperature comes close to zero. This is why this algorithm is also used as a provider of a warm point for the descent algorithm

[2] which is a near from the optimal quantizer of the uniform law on the unit square.

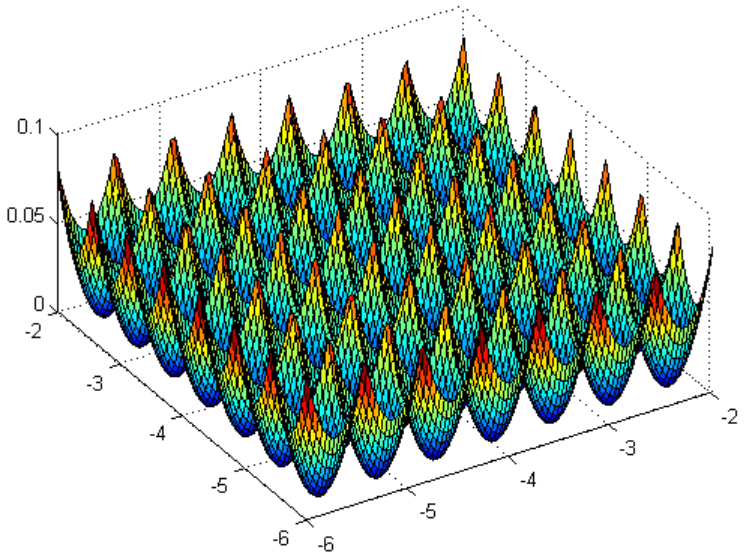

Figure 1.2: Numerical error: it is minimal on $\nabla\varphi^{-1}(y_i)$ where $\{y_i\}_{1\leq i\leq N}$ is the regular grid that quantizes the target measure $\mathcal{U}_{[-2,2]^2}$.

rather than used per se.

Quantization The quantization is used here to replace the evaluation of the expected value $\mathbf{E}(\varphi_v(X))$ and of the cell volume $\mu(V_j)$ by their discretized counterparts $\sum_{i=1}^N p_i\varphi_v(x_i)$ and $\sum_{i=1}^N p_i\mathbb{1}_{V_j}(x_i)$. Table 1.6 shows the results when the initial measure has 10000 atoms while the number of atoms of the target measure varies. It suggests that this method is able to achieve an error that is comparable to the exact descent, with a time complexity that has the same order of magnitude. However, unlike to the exact descent, both the computation time and the numerical error are not decreasing functions of the number of points, which makes it difficult to imply the actual time complexity and speed of convergence of this method. Experiments show that for a fixed number of points, the error can be lessened at the expense of refining the quantization grid of the initial measure. Yet, our experiments in this case do not suggest that this method should be preferred to the auction algorithm, and even less to the exact descent. This points out that when performing a descent method, the accurate computation of the gradient critically impacts the output of the algorithm.

Linear programming algorithms As said previously, the number of constraints increases very rapidly: for a grid of size N (i.e. N^2 points) on the square in $\mathbf{R}^2$, there are N^4 constraints. Practically, this makes the algorithm totally inefficient as soon as $N \geq 8$ in dimension 2. Experiments show that the simplex algorithm is ruled out, even for a small number of points, as it is very slow. The interior point method is more efficient, but very slow too. Furthermore

memory restrictions prevent from setting a number of points above 100 in dimension 2 due to the increasing number of constraints with N. Nevertheless, this algorithm provides a solution that has the same quality in terms of error than the one provided by the exact descent, but its time complexity is worse than the exact descent.

Table 1.1: Interior Point Algorithm, uniform case

Points number	CPU Time	Error
25	1.202	0.1602
36	5.806	0.1111
49	33.564	0.0816
64	285.358	0.0624
81	1038.1	0.0494

Descents algorithm combined with a warm point As every descent algorithm, the performance of the implemented BFGS procedure (the exact descent, or the quantization) is critically impacted by the quality of the starting point. Although some justifications are given above to choose a specific starting point, there is no insurance as to its 'optimality' in the case where μ and ν are not close. A starting point that works very well, i.e. a "warm point", is provided by running the auction algorithm or IPFP prior to run a descent algorithm. Figure 1.3 and table 1.3 show that a warm point speeds up consistently the convergence of the BFGS procedure (without improving the time-complexity yet, it just scales down the CPU time).

Uniform measure to normal measure

Exact descent can be still used, as the initial measure is uniform on the square, and we focus on this sole method. This case is not covered by our theoretical framework yet, as the target measure is not compact. Nevertheless, our experiments suggest that the same conclusions as for the first example hold. Figure 1.4 provides a comparison of the error achieved for a fixed number of points N, when the target measure is either optimally quantized or either approximated by a weighted sum of equally-weighted Dirac masses. The gain offered by quantization is obvious, as for a time complexity that is roughly comparable the numerical error is lower when using quantization.

Dilatation of a bivariate normal variable

A third type of example is investigated: it consists in the computation of the transport map between the standard bivariate normal law and a bivariate normal law with covariance matrix $2.Id$, so that both measures are not compactly supported anymore.
This is also the first case where the exact descent algorithm, that computes the volume of the V_i by triangulation can not be applied, as the initial measure is not the uniform one. The

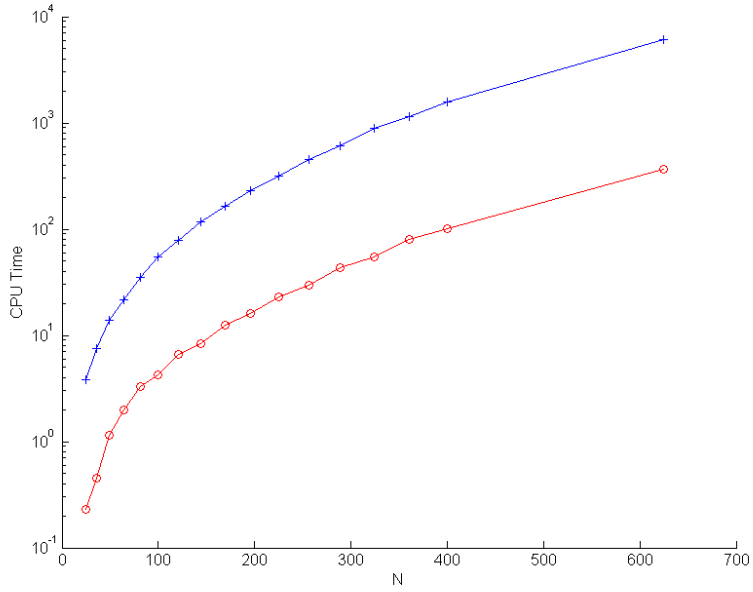

Figure 1.3: CPU time in seconds (log-scale) as a function of the number N of points for the exact descent algorithm applied to example 1. The curve with plus signs corresponds to no warm point; the curve with circles corresponds to an IPFP warm point (1000 points of discretization)

auction algorithm applied to a random discretization of both laws is compared to the descent algorithm using a 1000 points optimal quantization for the initial measure, coupled with IPFP as a warm point. The results are summarized in table 1.8 [3]. The conclusion remain similar to those drawn from the previous examples: IPFP is fast but generates error on the boundary that does not vanish when the discretization grid is refined.

The auction algorithm seems to be of less interest than the descent method, as it fails to provide an error below 10^{-1}, even with distributions that have more than 2000 atoms. Yet, it must be remarked that although the biggest quantization grid that we used to quantize the initial measure has 1000 points, this is not sufficient to obtain an uniform error that is below 10^{-2} with the descent algorithm.

A solution to this problem could be to implement an accurate method to compute the mass (here for the normal measure) of polytopes in order to apply the same technique as in the uniform case. In view of the result of the previous sections, it seems that non compactness is not such a big impediment to this method as is the need for an accurate computation of the objective function and its gradient.

[3] The definition of the numerical error must be adapted as the support is the whole plane. We use the fact that the bivariate standard normal law gives to the square $S_\alpha = [-q_{1-\alpha/2}, q_{1-\alpha/2}]^2$ a mass at least equal to $1-2\alpha$. α is set a t5% so that $S_{5\%} \approx [-1.96, 1.96]^2$. The numerical error is then defined as $\sup_{G \subset S_{2.5\%}} |\varphi_{opt} - \varphi_{num}|$ where G is a regular grid with 100×100 points on the square $S_{2.5\%}$.

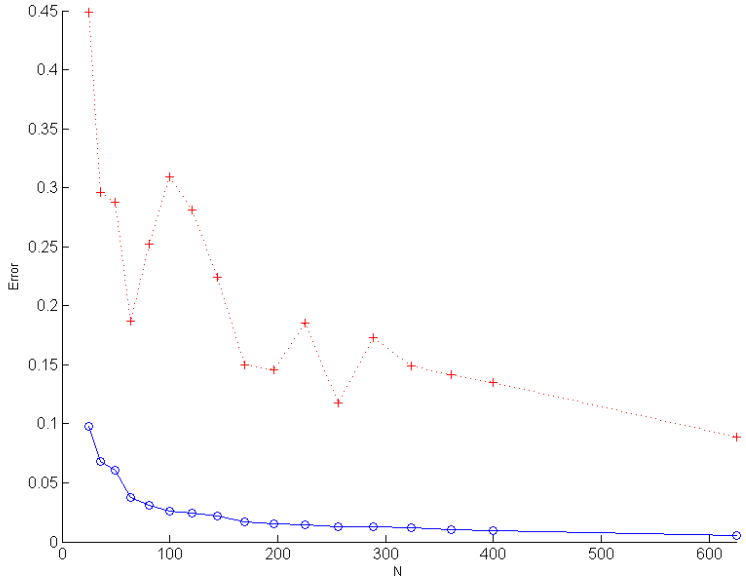

Figure 1.4: Example 2. Curve with circles: error as a function of the number of points N of the exact descent with quantization of the initial measure. Curve with crosses: error with a random discretization of the bivariate normal law.

1.5.5 Complexity and order of convergence in the planar case

Figure 1.6 displays the time complexity of the exact descent method for the two first examples. Experiments suggest that the complexity is $O(N^{5/2})$, at least in the case where the initial measure is uniform on the unit square. This is to be compared to the algorithm proposed in [3], that involves $O(N\log(N))$ operation at each iteration (but the number of iterations required for convergence is unknown). Recall the descent algorithm use a BFGS update, so that each step involves $O(N^2)$ operations; thus the complexity of this algorithm is necessarily bounded below by $O(N^2)$. Figure 1.7 displays the numerical error as a function of the number of points. Both examples suggest that the convergence has a rate of $O(1/\sqrt{N})$.

1.5.6 Convergence in higher dimension

In higher dimension, although the principle of the algorithm is still valid, the exact descent can become costly. Indeed, its complexity is driven by the speed at which Delaunay triangulations are computed. In dimension 2, these are done at a cost $O(N\log(N))$. As there are N cells, the gross computational cost of a single evaluation of the function $\mathcal{F}$ should be $O(N^2\log(N))$. In dimension $d>2$, algorithms run at the worst complexity $O(N^{\lceil\frac{d}{2}\rceil+1})$[4]. Hence, the time complexity of the exact descent is exponential with respect to the dimension.

[4] see for instance the DeWall algorithm, a divide and conquer algorithm presented in [8]

The auction algorithm seems at first sight less sensitive to the dimension. However, as suggested by the rate of convergence of the optimal quantization (c.f. theorem 2), when the dimension increases, the number of atoms that intervene in the optimal discretization of a continuous law of probability increases at an exponential rate. Thus, the curse of dimensionality is not overcome by any of the algorithms we tested.

1.6 Conclusion

This chapter showed both theoretically and empirically that the approximation of the optimal transport map can be done by first discretizing the target measure and then performing a descent algorithm. In particular in dimension 2 when the initial measure is uniform over a bounded polytope, this algorithm performs well compared to the auction algorithm. This method also benefits from the freedom left on the form of the discretization: we chose it 'optimal', in the sense of L^2 optimal quantization.
The same algorithm works when the dimension increases although the complexity is exponential with respect to the dimension. When the initial measure is not uniform over a polytope, it can be discretized. Yet we should emphasize that the performance of the descent algorithm suffers from such an approximation; this method best works when we have a means to compute accurately the mass of polytopes with respect to the initial measure.

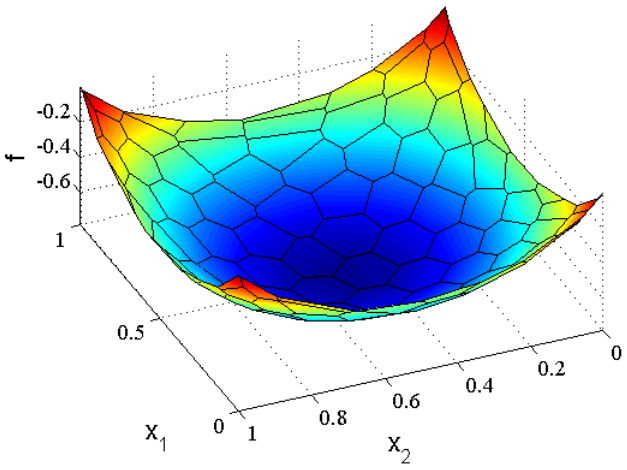

Figure 1.5: Second example: the computed Kantorovitch potential with a quantization grid of the normal law of size 100. The polytopes V_i appears on the surface plot.

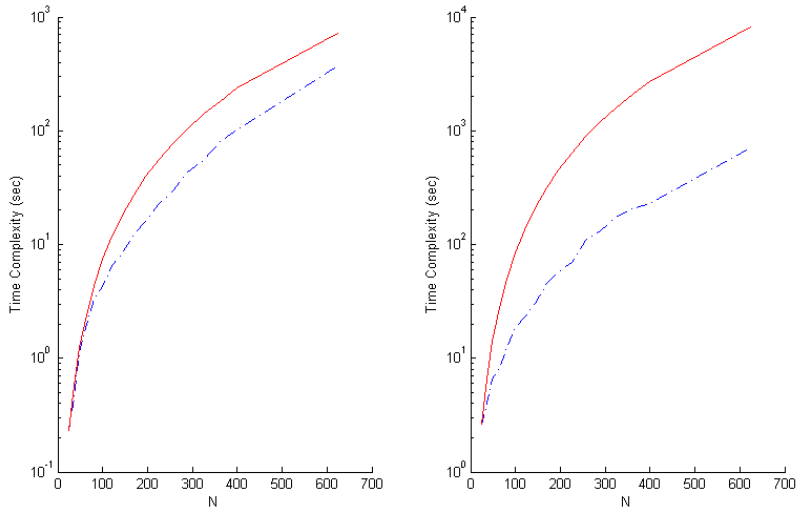

Figure 1.6: N is the number of points in the grid. Left: Example 1. Solid line is the curve $N^{5/2}$, the dashed line is the computed complexity. Right: the same for the second example

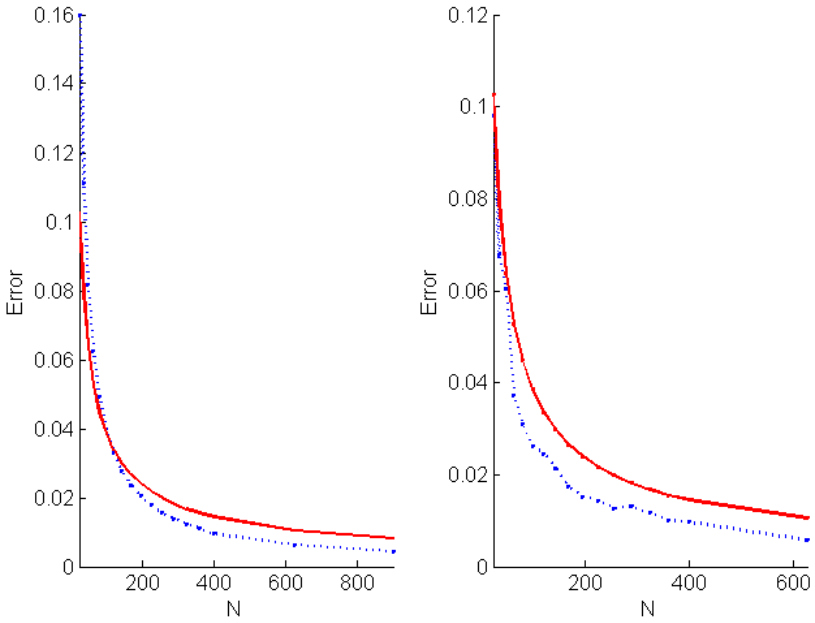

Figure 1.7: Left: Example 1. Right: Example 2. In both cases, the solid line is the curve $1/\sqrt{N}$ and the dashed line the computed error.

Bibliography

[1] S. Angenent, S. Haker, and A. Tannenbaum. Minimizing flows for the Monge-Kantorovich problem. *SIAM J. Math. Analysis*, 35:61–97, 2003.

[2] F. Aurenhammer, F. Hoffmann, and B. Aranov. Minkowski-type theorems and least square clustering. *Algorithmica*, 20:61–76, 1998.

[3] J.-D. Benamou and Y. Brenier. A numerical method for the optimal time-continuous mass transport problem and related problems. *Contemp. Math.*, 226:1–11, 1999.

[4] D. Bertsekas. Auction algorithms for network flow problems:a tutorial introduction. *Comput. Optim. Appl.*, 1, 1992.

[5] D. Bertsekas. Auction algorithm. Encyclopedia of Optimization, 2001.

[6] J. Borwein and A. Lewis. *Convex Analysis And Nonlinear Optimization, Theory and Examples*. Ouvrages de Mathématiques de la SMC. Springer, 2000.

[7] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.

[8] P. Cignoni, C. Montani, and R. Scopigno. Dewall: A fast divide & conquer Delaunay triangulation algorithm in e^d. *Computer Aided Design*, 30(5):333–341, 1997.

[9] A. Dominitz, S. Angenent, and A. Tannenbaum. On the computation of optimal transport maps using gradient flows and multiresolution analysis. In V. D. B. et al., editor, *Recent Advances in Learning and Control*, volume 371 of *LNCIS*, pages 65–78. Springer-Verlag Berlin Heidelberg, 2008.

[10] I. Ekeland, A. Galichon, and M. Henry. Comonotonic measures of multivariate risks. *Mathematical Finance*, 22:109–132, 2012.

[11] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.

[12] J. Kosowski and A. Yuille. The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Networks*, 7(3):477–490, 1994.

[13] J. J. Kosowsky and A. L. Yuille. The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Networks*, 7:477–490, 1994.

[14] M. Kvasnica, P. Grieder, and M. Baotić. Multi-Parametric Toolbox (MPT), 2004.

[15] G. Loeper and F. Rapetti. Numerical solution of the Monge-Ampère equation by a newton's algorithm. *C. R. Acad. Sci. Paris*, I 339, 2004.

[16] J. Nocedal and S. Wright. *Numerical Optimization, 2nd edition.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, 2006.

[17] G. Pagès and J. Printems. http://www.quantize.maths-fi.com.

[18] S. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol I and II.* Probability and Applications. Springer, 1998.

[19] L. Rüschendorf. Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, 23:1160–1174, 1995.

[20] L. Rüschendorf and L. Uckelmann. Numerical and analytical results for the transportation problem of Monge-Kantorovich. *Metrika*, 51:245–258, 2000.

[21] S.Graf and H.Luschgy. *Foundations of Quantization for Probability Distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer-Verlag, 2000.

[22] C. Villani. *Topics in Optimal Transportation.* Graduate Studies in Mathematics. American Mathematical Society, 2003.

[23] C. Villani. Stability of a 4-th order curvature condition arising in optimal transport theory. *J. Funct. Anal.*, 255(9):2683–2708, 2008.

[24] C. Villani. *Optimal Transport, Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, 2009.

1.7 Appendix

1.7.1 Proofs of various results

Proof of proposition 2

It consists in verifying the definition of a convex function. Let v and v' be in $\mathbf{R}^N$, and $t \in [0,1]$. For all $x \in \mathbf{R}^n$,

$$\begin{aligned}(1-t)\varphi_v(x) + t\varphi_{v'}(x) &= (1-t)\max_i[x \cdot y_i - v_i]) + t\max_i[x \cdot y_i - v'_i]) \\ &\geq (1-t)(x \cdot y_i - v_i) + t(x \cdot y_i - v'_i), \text{ for all } i \\ &= x \cdot y_i - ((1-t)v_i + tv'_i)\end{aligned}$$

Thus, for every x, $(1-t)\varphi_v(x) + t\varphi_{v'}(x) \geq \max_i[x \cdot y_i - ((1-t)v_i + tv'_i)] = \varphi_{(1-t)v+tv'}(x)$. The convexity follows by taking the expected value.
The function is bounded below: let $v \in \mathbf{R}^n$ and $\tilde{v} = v - v_{i_0}\mathbb{1}$ where $i_0 \in \operatorname{argmin}_i v_i$ so that $\tilde{v}$ has nonnegative components and $\min_i \tilde{v}_i = 0$.

$$\mathcal{F}(v) = \mathcal{F}(\tilde{v}) = \int \varphi_v(x)d\mu(x) + q'v \geq y_{i_0} \cdot \int x d\mu - \underbrace{\tilde{v}_{i_0}}_{=0} + \underbrace{q'\tilde{v}}_{\geq 0} \geq -(\max_j |y_j|)|\mathbf{E}(X)|$$

The less obvious point that remains to prove is actually the smoothness of $\mathcal{F}$ and the form of its gradient, although the formula is easily found formally. We prove the existence and the continuity of the directional derivatives with respect to the canonical basis $e_j, j = 1, \ldots, N$ of the function $v \mapsto \int \varphi_v(x)d\mu(x)$. First, for a given x, $\lim_{t\downarrow 0} \frac{\varphi_{v+te_j}(x) - \varphi_v(x)}{t}$ exists is equal to $-\mathbb{1}_{V_i}(x)$, by an application of the envelope theorem, see prop 2.3.2 in [6].
As, for all x, $|\varphi_{v+te_j}(x) - \varphi_v(x)| \leq |t|$, the fraction $\frac{\varphi_{v+te_j}(x) - \varphi_v(x)}{t}$ is uniformly bounded with respect to t, and one can invert limit and integral. The same applies to $\lim_{t\uparrow 0} \frac{\varphi_{v+te_j}(x) - \varphi_v(x)}{t}$, except that it is equal to $-\mathbb{1}_{Int(V_j)}(x)$ and therefore,

$$\lim_{t\uparrow 0} \int \frac{\varphi_{v+te_j}(x) - \varphi_v(x)}{t} d\mu(x) = -\mu(Int(V_i))$$

However, as ∂V_i has Lebesgue measure zero, and μ is supposed to not charge small sets, this equals $\mu(V_i)$. Eventually, $\mathbf{E}(\varphi_v(X))$ has partial derivatives with respect to v_i for all i, viz. $v \mapsto \mu(V_i(v))$ which are continuous functions.□

Proof of lemma 1

First, every v_i^k can be chosen nonnegative as $(v_i^k)_{1 \leq k \leq N_k}$ is defined up to a multiple of $(1, \ldots, 1)$. For example one can impose for all k, $\min_{i=1\ldots k} v_i^k = 0$. For each k and $1 \leq i \leq k$, let $x_{i,k}$ be in V_i^k (such a point exists as $\mu(V_i^k) = q_k > 0$). Moreover, let j_k such that $v_{j_k}^k = 0$. The

definition of $x_{i,k}$ being in V_i^k implies:

$$x_{i,k} \cdot y_i^k - v_i^k \geq x_{i,k} \cdot y_{j_k}^k$$

which gives

$$v_i^k \leq x_{i,k} \cdot (y_i^k - y_{j_k}^k) \leq \sup_{\text{supp}\mu} |x|.diam(\text{supp}\nu) < +\infty$$

Proof of lemma 2

We show that $\{\varphi_k\}$ satisfies the requirements of the Ascoli theorem, namely that it is pointwise bounded and equicontinuous. It is pointwise bounded, as the sequence is uniformly bounded on suppμ. The equicontinuity results from the fact that this sequence is equi-Lipschitz, i.e. each φ is Lipschitz with a Lipschitz constant that does not depend on k. Indeed, if x, x' belong to suppμ, and $x \neq x'$:

$$\begin{aligned}
\varphi_k(x) - \varphi_k(x') &= \max_i(x \cdot y_i^k - v_i^k) - \max_i(x' \cdot y_i^k - v_i^k) \\
&= x \cdot y_{i(x)}^k - v_{i(x)}^k - \max_i(x' \cdot y_i^k - v_i^k) \\
&\leq x \cdot y_{i(x)}^k - v_{i(x)}^k - x' \cdot y_{i(x)}^k + v_{i(x)}^k \\
&= (x - x') \cdot y_{i(x)}^k \leq |x - x'| \sup_{y \in \text{supp}\nu} |y|
\end{aligned}$$

Similarly, $\varphi_k(x) - \varphi_k(x') \geq -|x - x'| \sup_{y \in \text{supp}\nu} |y|$, and consequently $|\varphi_k(x) - \varphi_k(x')| \leq |x - x'|$.

1.7.2 Numerical Results

Table 1.2: Exact descent, example 1

Point Number	CPU Time (sec)	$\|\| \cdot \|\|_{+\infty}$ Error
25	3.822	0.1600
64	21.512	0.0624
81	34.866	0.0494
100	55.287	0.0400
196	230.428	0.0204
225	316.464	0.0178
256	457.518	0.0156
361	1147	0.0111
400	1579	0.0096
625	6148.9	0.0065

Table 1.3: Exact descent with IPFP warm point, example 1

Point Number	CPU Time (sec)	$\|\| \cdot \|\|_{+\infty}$ Error
25	0.23	1.60E-01
64	1.99	6.24E-02
81	3.29	4.94E-02
100	4.27	4.00E-02
196	15.94	2.04E-02
225	23.15	1.78E-02
256	29.50	1.56E-02
361	80.28	1.11E-02
400	100.91	9.60E-03
625	368.61	6.40E-03
900	1346.20	4.50E-03

Table 1.4: Auction algorithm, example 1

Point Number	CPU Time	$\|\|\cdot\|\|_{+\infty}$ Error
25	0.03	1.43E+00
100	1.06	5.83E-01
225	5.70	3.79E-01
625	50.64	9.88E-02
900	113.68	5.36E-02
1225	248.32	6.85E-02
1600	467.93	3.23E-02
2025	782.34	3.36E-02
2500	1288.70	4.78E-02
3600	3387.00	1.87E-02
4900	6174.00	1.89E-02
6400	14834.00	1.97E-02

Table 1.5: IPFP algorithm, example 1

Point Number	CPU Time	$\|\|\cdot\|\|_{+\infty}$ Error
25	1.60E-02	1.60E-01
81	3.10E-02	4.57E-02
100	3.10E-02	4.68E-02
169	9.68E-01	3.90E-02
196	9.40E-02	3.55E-02
225	1.10E-01	1.78E-02
256	1.25E-01	3.43E-02
289	1.40E-01	2.01E-02
400	1.88E-01	3.35E-02
625	3.28E-01	2.85E-02
900	6.41E-01	3.60E-02
1024	7.50E-01	3.65E-02

Table 1.6: Descent algorithm, 10000 points quantization grid for μ, example 1

Point Number	CPU Time	$\|\|\cdot\|\|_{+\infty}$ Error
25	9.38	1.82E-01
36	46.10	1.20E-01
100	28.48	5.68E-02
121	153.97	3.27E-02
196	178.94	2.01E-02
225	151.72	2.64E-02
256	233.61	1.79E-02
361	633.99	1.21E-02
400	108.86	1.48E-02
784	1308.50	5.70E-02
900	729.50	6.70E-03
1024	697.10	9.60E-03
1225	1133.00	1.36E-02

Table 1.7: Exact descent, example 2

Point Number	CPU Time	$\|\|\cdot\|\|_{+\infty}$ Error
25	2.635	9.82E-02
64	8.131	3.72E-02
81	12.474	3.10E-02
100	18.515	2.61E-02
196	58.027	1.52E-02
225	69.814	1.43E-02
256	109.037	1.26E-02
289	132.491	1.31E-02
361	207.583	1.01E-02
400	230.204	9.70E-03
625	712.705	5.70E-03

Table 1.8: Example 3, results

	Point Number	CPU Time	Error
Descent	25	4,22	5,10E-01
	100	12.40	1.61E-01
	200	27.18	6.36E-02
	300	125.49	9.15E-02
	400	114.86	8.47E-02
	500	130.06	5.88E-02
	600	386.67	4.15E-02
	700	498.63	6.44E-02
	800	591.36	4.52E-02
	900	458.02	6.70E-02
Auction	25	0.02	4.13E+00
	100	0.19	1.15E+00
	500	4.33	6.64E-01
	1000	33.20	3.74E-01
	2000	153.70	2.29E-01
	2500	357.03	1.24E-01
	3000	352.53	1.37E-01
	5000	840.42	2.43E-01
IPFP	25	< 0.01	5.24E-01
	100	0.02	1.93E-01
	300	0.06	7.83E-02
	500	0.17	5.09E-02
	700	0.27	3.46E-02
	900	0.36	3.02E-02
	1000	0.41	2.72E-02

Chapter 2

Extreme dependence for multivariate data [1]

Introduction

Extreme dependence, and the closely related notion of comonotonicity are important concepts in various fields. It is central in the economics of insurance (following the seminal work of Borch [3] and Arrow [1], [2]), in economic theory (see [21], [12] and [19]), and in statistics (see [5], [17], [16], [22]).

Extreme positive dependence between two real random variables (X, Y) is characterized by the fact that their cumulative distribution function should satisfy $F_{X,Y}(x, y) = \min(F_X(x), F_Y(y))$, or equivalently, that their copula C should be the upper Fréchet copula $C(u_1, u_2) = min(u_1, u_2)$. This form of dependence occurs when X and Y are comonotone, i.e. when both X and Y can be written as nondecreasing functions of a third random variable Z (for instance one may choose $Z = X + Y$). As a consequence, comonotone variables maximize covariance over the set of pairs with fixed marginals:

$$\mathbf{E}(XY) = \sup_{\substack{\tilde{X}\sim X \\ \tilde{Y}\sim Y}} \mathbf{E}(\tilde{X}\tilde{Y}), \tag{2.1}$$

where $\tilde{X} \sim X$ denotes the equality in distribution between $\tilde{X}$ and X. Similarly, X and Y are said to have extreme negative dependence when X and $-Y$ have extreme positive dependence. Their covariance is then minimized instead of maximized, and their copula is now the lower Fréchet copula $C(u, v) = \max(u + v - 1, 0)$.

The present chapter aims at proposing an operational theory of extreme dependence in the multivariate case, namely when X and Y are random vectors. Our contribution is twofold.

[1] This chapter is a joint work with Alfred Galichon.

First, we shall introduce a generalization of the notion of extreme dependence to the multivariate case, and we investigate how extreme positive dependence generalizes in this setting. Then we shall apply these ideas in a financial context to stress-testing dependence, i.e. we shall investigate the sensitivity of a portfolio on the strength of dependence between two random vectors.

Generalizing extreme dependence. When dealing with the multivariate case, where X and Y are random vectors in $\mathbb{R}^d$, there is no canonical way to generalize this notion of (positive or negative) extreme dependence and Fréchet copula. One first approach, based on the theory of Optimal Transport (see eg. [20]) would be to consider the following optimization problem

$$\max_{\substack{\tilde{X}\sim X\\ \tilde{Y}\sim Y}} \mathbf{E}(\tilde{X}\cdot\tilde{Y}) \tag{2.2}$$

where $\cdot$ is the scalar product in $\mathbb{R}^d$. This program is a multivariate extension of the covariance maximization problem (2.1) and defines as extreme the distribution of the pair $(\tilde{X},\tilde{Y})$ solution to the above problem However this does not take into account the cross-dependence between X_i and Y_j for $i\neq j$, and therefore seems quite arbitrary for our purposes.

A more satisfactory generalization is based on the idea that both positive and negative extreme dependence are obtained by the maximization of a nonzero bilinear form in (X,Y) over the set of couplings of X and Y (i.e. joint distributions with fixed marginals). That is, consider solutions of (2.2) where the scalar product is replaced by any nonzero bilinear form. This will be our notion of *multivariate extreme dependence*: random vectors X and Y shall exhibit extreme dependence if their cross-covariance matrix maximizes the expected value of a nonzero bilinear form over all the couplings of X and Y. These extreme coupling are proposed as a generalization of Fréchet (positive and negative) extreme dependence in the multivariate case. We shall provide a natural geometric characterization of this notion by considering the *covariogram* , that is the set of all cross-covariance matrices $\mathbf{E}(XY')$ for all the couplings of X and Y. Then X and Y have extreme dependence if and only if their cross-covariance matrix lies on the boundary of the covariogram.

We then turn to generalizing the notion of extreme *positive* dependence. One natural way to generalize extreme positive dependence is to look for the couplings $(\tilde{X},\tilde{Y})$ that yield a cross-covariance matrix $Cov(\tilde{X},\tilde{Y})=\mathbb{E}(\tilde{X}\tilde{Y}')=(\mathbb{E}(\tilde{X}_i\tilde{Y}_j))_{i,j}$ which would be maximal elements for a certain partial (conical) ordering on matrices. As we shall see, it turns out that under this definition, extreme positive dependence implies extreme dependence, and we can characterize the geometric locus of extreme positive dependent vectors on the covariogram.

Stress-testing dependence. We give a method to associate any coupling, for example any empirical coupling, with an extreme coupling, by means of entropic relaxation technique. An

algorithm is described and results concerning its implementation are given. Applications to financial data are provided, yielding the definition of indices of maximal correlation as well as a prospective application to progressive stress-tests of dependence.

The chapter is organized as follows: the first section presents the notion of covariogram and the definition of couplings with extreme dependence deduced thereof, as well a characterization of such couplings. The second section defines couplings with positive extreme dependence; a characterization of these couplings makes the connection with the notion of extreme dependence. The third section defines an index of dependence, the affinity matrix; a method to associate any coupling with an extreme coupling is described. We conclude with financial applications, namely stress-testing portfolio allocations and options pricing, as well as the computation of indices with extreme dependence. All proofs are collected in 2.9.

Notations, definitions. We make the following distinction between the *univariate* case and the *multivariate* case. We refer to the univariate case when considering the dependence between real valued random variables: this is the subject of the theory of copulas. In most of this chapter we consider *random vectors*, and the dependence between two random vectors; in this case we speak of *multivariate dependence*.

Let P and Q be two probability laws on $\mathbb{R}^I$ and $\mathbb{R}^J$, with finite second order moments. Without restricting the generality we assume that P and Q have null first moments, so that the second order moments $\mathbf{E}(X_i Y_j)$ are indeed covariances. $\Pi(P,Q)$ is the set of all probability laws over $\mathbb{R}^I \times \mathbb{R}^J$ having marginals P and Q. We refer to an element of $\Pi(P,Q)$ as a *coupling*, understating the probabilities P and Q. If M and N belong to $\mathbf{M}_{I,J}(\mathbb{R})$, their scalar product is denoted by $M \cdot N = Tr\,(M'N)$. If $(X,Y) \sim \pi \in \Pi(P,Q)$, we denote indifferently $\sigma_{X,Y}$ or σ_π the matrix with general term $\mathbf{E}(X_i Y_j)$, which is the covariance between X_i and Y_j; it is the cross-covariance matrix between X and Y. Remark that $\sigma_{X,Y}$ is the upper-right block of the variance-covariance matrix of the vector $Z = (X,Y)'$, and that $\sigma_{X,Y}$ is not symmetric in general.

Eventually, let us recall that the *subdifferential* $\partial f(x_0)$ of a convex function on $\mathbb{R}^n$ at a point x_0 is defined as set of vectors v such that $f(x) - f(x_0) \geq v \cdot (x - x_0)$ for all $x \in \mathbf{R}^n$. Here the dot is the usual scalar product. It reduces to $\{\nabla f(x_0)\}$ if f is differentiable at x_0, which is true for almost every x_0 according to Rademacher theorem.

2.1 Related literature and contribution

As mentioned in the introduction, the extension to the multivariate setting of the correlation maximization problem (2.1) has been tackled by several authors, especially to define notions of *multivariate comonotonicity*. Puccetti and Scarsini [15] list several possible definitions of

multivariate comonotonicity, among which two of them are directly related to the variatonal problem (2.2). Namely, *c-comonotonicity* refers to the couplings that are by solving problem (2.2): these are the optimal quadratic couplings of Optimal Transport Theory. This variational approach to multivariate comonotonicity is also the basis of Ekeland, Galichon and Henry [7] and Galichon and Henry [9]. They propose to extend the univariate notion of comonotonicity and define the *μ-comonotonicity* by stating that two vectors X and Y are μ-comonotone if there exists a random vector $U \sim \mu$ such that

$$\begin{aligned} \mathbf{E}(X \cdot U) &= \max\{\mathbf{E}(X \cdot \tilde{U}),\, \tilde{U} \sim \mu\} \\ \mathbf{E}(Y \cdot U) &= \max\{\mathbf{E}(Y \cdot \tilde{U}),\, \tilde{U} \sim \mu\} \end{aligned}$$

This notion of comonotonicity has the advantage of being transitive, unlike c-comonotonicity. Carlier, Dana and Galichon [4] showed that this notion of comonotonicity appeared as 'more natural' than the other ones because it is directly related to Pareto efficiency.

This chapter aims at finding multivariate couplings which exhibit a form of strong dependence, just as the previously defined comonotonic couplings. The couplings that are defined as 'extreme' in what follows, are comonotonic couplings (in the sense of the c-comonotonicity) *up to a linear transform* of one marginal (the c-comotonic coupling corresponds to the identity transform). In other words, an extreme coupling (X, Y) satisfies the variational problem

$$\mathbf{E}(X'MY) = \sup_{\pi \in \Pi(P,Q)} \mathbf{E}_\pi(X'MY) \tag{2.3}$$

This definition of extreme dependence is broad enough to encompass 'positive dependence' as c-comonotonicity as well as 'negative dependence' (counter-comonotonicity in the univariate case). Furthermore, it allows for a geometrical interpretation of extreme dependence: an extreme coupling has a cross-covariance matrix located on the boundary of the compact and convex set of all possible cross-covariance matrices, called the covariogram. This set has been introduced in Galichon and Salanié [10], who point out the importance of its boundary. Taking advantage of this simple interpretation, we then investigate the couplings π which have cross-covariance matrix σ_π that are maximal for some partial orders $\succ$, and show that they form an easily characterized subset of the extreme dependent couplings. The rest of the chapter consists in computing the extreme couplings, and, for any given coupling $\hat{\pi}$ propose a means to build a continuous sequence of couplings π_T with π_0 being extreme, and $\sigma_{\pi_1} = \sigma_{\hat{\pi}}$. This is done by penalizing the problem (2.3) with an entropy term, which allows for fast computations when the marginals are discrete law of probability, thanks to the Iterative Proportional Fitting Procedure. This algorithm dates back to Deming and Stephan (1940) [6], and has been used by Yuille and Kosowosky [11] (although they do not refer explicitly to IPFP, their method is equivalent to it) to solve the assignment problem, and in Econometrics in [10].

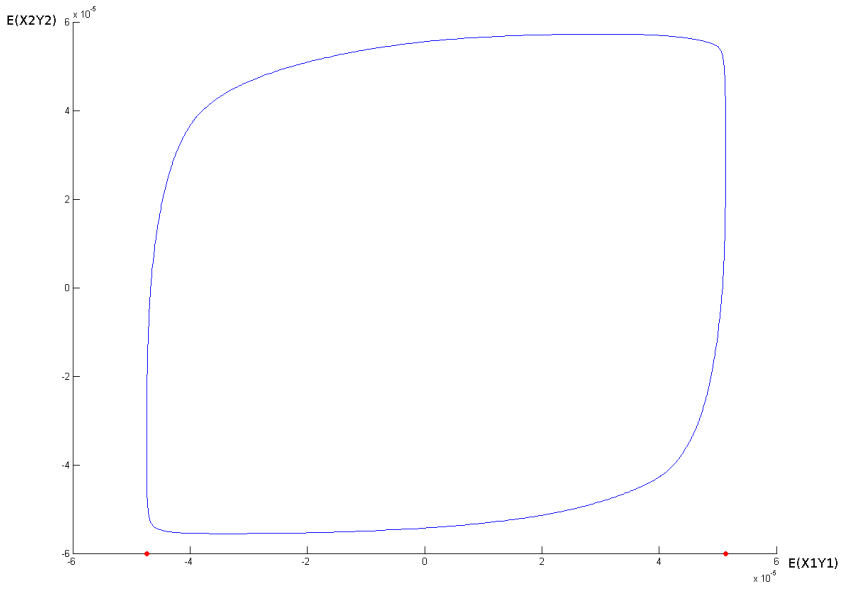

Figure 2.1: Example of a 2 dimensional section of a covariogram

2.2 Multivariate extreme dependence

In this section we describe our proposed notion of multivariate extreme dependence. Consider the covariogram, the set of cross-covariance matrix $(\mathbf{E}_\pi(X_i Y_j))_{i,j}$ in $\mathbf{M}_{I,J}(\mathbb{R})$ for any $\pi \in \Pi(P,Q)$:

Definition 1 *The covariogram $\mathcal{F}(P,Q)$ is defined by:*

$$\mathcal{F}(P,Q) = \{\Sigma \in \mathbf{M}_{I,J}(\mathbb{R}) : \exists \pi \in \Pi(P,Q), \Sigma_{ij} = \mathbf{E}_\pi(X_i Y_j), \textit{for all } i,j\}\,.$$

As $\Pi(P,Q)$ is a convex and compact set (a proof of this last property can be found in [20], pp. 49-50) the covariogram is itself a convex compact subset of $\mathbf{M}_{I,J}(\mathbb{R})$.

Figure 2.1 gives a first example of the 2 dimensional section of a covariogram where only the diagonal elements of the cross-covariance matrix are represented, when $I = J = 2$. P and Q are discrete distributions on $\mathbb{R}^2$ with equally weighted atoms and we look at the two first component-wise covariances $\mathbf{E}(X_1 Y_1)$, $\mathbf{E}(X_2 Y_2)$. The solid curve is the boundary of the covariogram: every coupling between P and Q would have a cross-covariance matrix that projects within the convex hull of this curve. The independence coupling projects on the point $(0,0)$. The dots on the x-axis represent respectively the minimal and maximal covariances between X_1 and Y_1. These covariances would be attained in the copula framework by the lower and upper Fréchet copulas. This motivates our definition of *extreme dependence couplings* as couplings whose projection lies on the boundary of the covariogram.

Definition 2 *A coupling* $(X,Y) \sim \pi \in \Pi(P,Q)$ *has extreme dependence if and only if* $(\mathbf{E}_\pi(X_iY_j))_{ij}$ *lies on the boundary of the covariogram* $\mathcal{F}(P,Q)$.

The cross-covariance matrix between X and Y, $\sigma_{X,Y}$, enjoys the property

$$Tr\left(M'\sigma_{X,Y}\right) = \mathbf{E}(X'MY), \text{ for all } M \in \mathbf{M}_{I,J}(\mathbb{R}) \tag{2.4}$$

which allows us to reformulate the notion of extreme dependence as follows:

Theorem 3 *The following conditions are equivalent:*

i) $(X,Y) \sim \pi \in \Pi(P,Q)$ *have extreme dependence;*

ii) there exists $M \in \mathbf{M}_{I,J}(\mathbb{R})\backslash\{0\}$ *such that*

$$Tr\left(M'\sigma_\pi\right) = \sup_{\tilde{\pi}\in\Pi(P,Q)} Tr\left(M'\sigma_{\tilde{\pi}}\right)$$

or equivalently

$$\mathbf{E}_\pi(X'MY) = \sup_{\tilde{\pi}\in\Pi(P,Q)} \mathbf{E}_{\tilde{\pi}}(X'MY); \tag{2.5}$$

iii) there exists $M \in \mathbf{M}_{I,J}(\mathbb{R})\backslash\{0\}$ *and a convex function* u *on* $\mathbb{R}^I$ *such that* $M.Y \in \partial u(X)$ *holds almost surely.*

In dimension 1, the interpretation is obvious: two real random variables have extreme dependence iff there exists a scalar $m \neq 0$ and a nondecreasing function u such that $mY = u(X)$. According to the classic terminology, X and Y are said comonotonic if $m > 0$, and anti-comonotonic otherwise.
When $M = Id$ in (2.5), the optimal coupling is the optimal transport coupling for the quadratic cost solving (2.2).

2.3 Positive extreme dependence

The aim of this section is to propose a generalization of the concept of Fréchet copula of upper dependence to the multivariate case. As already mentioned, copula theory fails to handle this problem. Indeed, if C_P and C_Q are two copulas, the first one of order I (associated with distribution P) and the second of order J (associated with distribution Q), a natural candidate for being the copula of positive extreme dependence would be $C_\pi(x,y) = \min(C_P(x_1,\ldots,x_I), C_Q(x_1,...,x_J))$. But according to the so-called 'Impossibility theorem' (see [14]) , C_π is a copula function if and only if C_P and C_Q are respectively the upper Fréchet copula of order I and J. We thus depart from the copula approach and aim at characterizing positive extreme dependence directly through the cross-covariance matrix of X and Y. Starting from the simple observation that in the univariate case, the positive extreme

dependence attains maximum covariance between X and Y over all the couplings of P and Q, we shall introduce a conic order on the cross-covariance matrices $\sigma_{X,Y}$ and define positive extreme dependent couplings as the couplings whose cross-covariance matrix is a maximal element for that order.

In what follows one considers convex cones that are used to define conic orders . In order for our results to hold, they are assumed to have a particular form, namely dual cones of cones with compact basis (2.8 provides some background on such cones). More precisely, for each compact convex set $C \subset \mathbf{M}_{I,J}(\mathbb{R})$ such that $0 \notin C$ (such a set is called a compact basis), a closed convex cone in $\mathbf{M}_{I,J}(\mathbb{R})$ is defined by setting:

$$K(C) = \{y \in \mathbf{M}_{I,J}(\mathbb{R}) | x \cdot y \geq 0, \, \forall x \in C\} \tag{2.6}$$

Considering cones of this form might seem restrictive, but we provide examples that show that many classic cones can be defined in such a manner.

$K(C)$ defines a conic order on $\mathbf{M}_{I,J}(\mathbb{R})$. More precisely, a *strict* conic order is needed and we set, for A and B two matrices in $\mathbf{M}_{I,J}(\mathbb{R})$

$$A \succ_{K(C)} B \text{ if } A - B \in Int(K(C))$$

The interior of $K(C)$ is $\{y \in \mathbf{M}_{I,J}(\mathbb{R}) | x \cdot y > 0, \, \forall x \in C\}$. Let $K = K(C)$ be such a cone.

Definition 3 *A coupling (X, Y) such that $\sigma_{X,Y}$ is a maximal element in $\mathcal{F}(P, Q)$ with respect to the strict conic order $\succ_K$ is said to have* positive extreme dependence *with respect to $\succ_K$.*

The following results fully characterize couplings with positive extreme dependence in terms of maximal correlation couplings.

Theorem 4 *The following conditions are equivalent:*

i) $(X, Y) \sim \pi \in \Pi(P, Q)$ have extreme positive dependence with respect to $\succ_K$;

ii) there exists $M \in C$ such that

$$Tr\left(M'\sigma_\pi\right) = \sup_{\tilde{\pi} \in \Pi(P,Q)} Tr\left(M'\sigma_{\tilde{\pi}}\right)$$

or equivalently

$$\mathbf{E}_\pi(X'MY) = \sup_{\tilde{\pi} \in \Pi(P,Q)} \mathbf{E}_{\tilde{\pi}}(X'MY); \tag{2.7}$$

iii) there exists $M \in C$ and a convex function u such that $M.Y \in \partial u(X)$ holds almost surely.

Hence, $\sigma_{X,Y}$ is maximal if and only if there exists $M \in C$ such that X and MY are maximally correlated for the scalar product. Obviously, this result is a close parallel to Theorem 3 except

that M is constrained to belong to C. As a consequence the positive extreme couplings are a particular case of extreme couplings. Once again the interpretation in dimension 1 is straightforward: X and Y have positive extreme dependence (w.r.t. the order in $\mathbf{R}$) iff they are comonotonic.

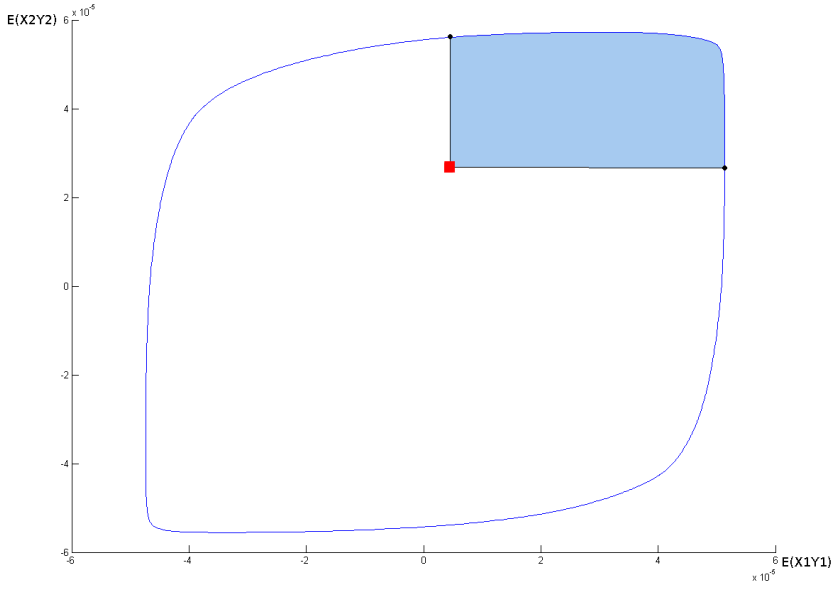

Figure 2.2: Shaded region: location of the couplings dominating the coupling materialized by the square dot.

To better understand the relation between those two types of couplings, let us go back to the two dimensional section of the covariogram discussed in the previous section, and take for K the positive orthant of $\mathbb{R}^2 \times \mathbb{R}^2$. The shaded region in Figure 2.2 is the set of couplings dominating the coupling that projects on the square dot, with respect to that order; as a consequence this coupling can not have positive extreme dependence. This intuitively explains why maximal elements should be on the boundary of the covariogram, hence that positive extreme couplings should be extreme couplings. Maximal elements are represented on the bold line figure 2.3: those are not dominated by an element of the covariogram. Consequently the couplings exhibiting positive extreme dependence, i.e. the one than can not be dominated, are located as shown in Figure 2.3. They are on the bold portion of the boundary, in the upper right corner of the covariogram, and forms only a 'small' part of the couplings of extreme dependence.

To demonstrate the applicability of this approach, we now give several examples of partial orders on covariance matrices.

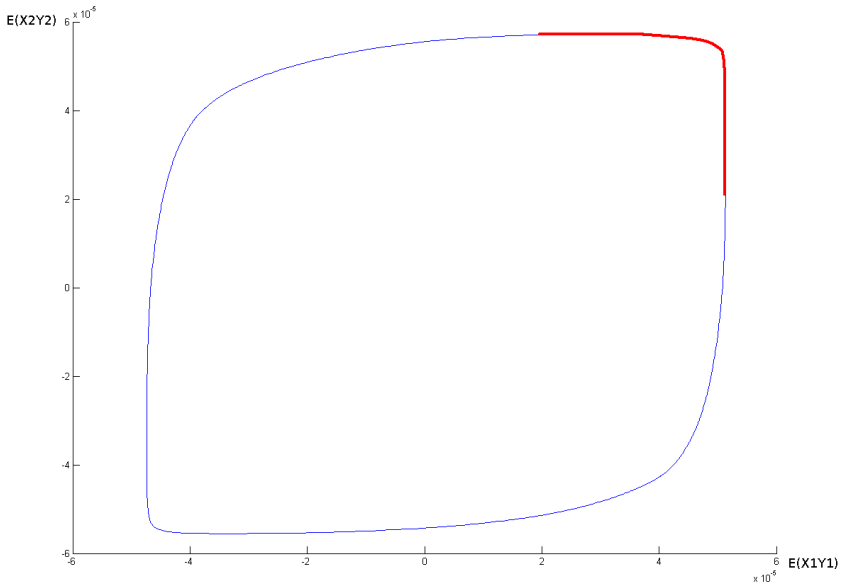

Figure 2.3: Maximal couplings on the boundary.

Example 1 (Orthant order) *Let $M^+_{I,J}(\mathbb{R})$ (resp. $M^{++}_{I,J}(\mathbb{R})$) denotes the set of real $I \times J$ matrices with nonnegative coefficients (resp. positive coefficients). The set $C = M^+_{I,J}(\mathbb{R}) \cap \{\sum_{i,j} M_{i,j} = 1\}$ is a compact basis set. $K(C)$ is easily seen to be the set $M^+_{I,J}(\mathbb{R})$ and its interior is $M^{++}_{I,J}(\mathbb{R})$. Eventually $A \succ B$ iff $A - B$ has only positive coefficients: this is the (strict)* orthant order *on matrices.*

Example 2 (Loewner order) *Let S^+_n and S^{++}_n denote respectively the set of nonnegative matrices in S_n and the set of definite positive matrices in S_n. If $C = \{S \in S^+_n(\mathbb{R}) | Tr(S) = 1\}$ is the set of semi-definite matrices with unit trace, C is a convex compact subset of $\mathbf{M}_n(\mathbb{R})$ and $K(C) = \{M \in \mathbf{M}_n(\mathbb{R}) | Tr(M'S) \geq 0, \forall S \in C\}$ is the set of matrices M whose symmetric part, $\frac{M+M'}{2}$, is semi-definite positive. The strict order $\succ_{K(C)}$ is then defined as: $A \succ B$ iff the symmetric part of $A - B$ is definite positive. This is an extension to $\mathbf{M}_n(\mathbb{R})$ of the classic* Loewner order *on symmetric matrices.*

The following trivial example shows that this ordering allows various positive extreme couplings. A first remark is that the maximum correlation coupling is indeed positive extreme, by setting $M = Id$ in theorem 4. Consider $P \sim \mathcal{N}(0, I_2)$, the bivariate normal law, and $Q = \mathcal{N}(0,1) \otimes \mathcal{U}_{(0,1)}$, the law of a vector whose first component is normal and the second one is the uniform law on $(0,1)$, independent from the first component. Let $X \sim P$ and $Y = (X_1, U)'$, $U \sim \mathcal{U}_{(0,1)}$ independent from (X_1, X_2), so that $Y \sim Q$. This coupling has not the maximum correlation even though $X_1 = Y_1$. However it satisfies (2.7) with $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and can be qualified as a maximal coupling.

Example 3 (Hermitian order) *Let*

$$M_S = \frac{M + M'}{2} \quad , \quad M_A = \frac{M - M'}{2}$$

the symmetric and skew-symmetric part of a matrix $M \in M_n(\mathbb{R})$. *We define the* hermitian transform $\tilde{M} \in M_n(\mathbb{C})$ *of* M *by setting*

$$\tilde{M} := M_S + iM_A, \text{ where } i^2 = -1$$

As iA *is hermitian as soon as* A *is skew-symmetric,* $\tilde{M}$ *is hermitian. Using the Loewner order on hermitian matrices we define a (partial) strict order on* $M_n(\mathbb{R})$ *by setting*

$$M \succ 0 \overset{def}{\Leftrightarrow} \tilde{M} \succ 0$$

If $C = \{M \in M_n(\mathbb{R}) | \tilde{M} \in S_n^+(\mathbb{C}), Tr(M) = 1\}$, *then* $K(C) = \{M \in M_n(\mathbb{R}) | \tilde{M} \in S_n^+(\mathbb{C})\}$.

2.4 An index of dependence

Suppose now we are observing or simulating a coupling $\hat{\pi} \in \Pi(P, Q)$, for instance an empirical coupling. Even if this coupling is supposed to exhibit strong dependence, its cross-covariance matrix will never be exactly located on the boundary of the covariogram. Our problem is then to *associate an extreme coupling with* $\hat{\pi}$; more precisely, we propose to find a continuous sequence of non deterministic couplings π_T such that $\pi_1 = \hat{\pi}$ and π_0 is an extreme coupling. In other words, we give a means to go smoothly from an empirical coupling to an extreme one by progressively increasing the strength of the dependence between the marginals. This is done by introducing an entropic penalization of (2.5), so that its solutions project on inner points of the covariogram.

2.4.1 Entropic relaxation

We introduce temperature in (2.5) by means of an entropy term ; it becomes

$$W(M, T) := \max_{\pi \in \Pi(P,Q)} \left(\mathbf{E}_\pi(X'MY) + TEnt(\pi) \right) \tag{2.8}$$

The entropy of a coupling π is defined as

$$Ent(\pi) = \begin{cases} -\int \log \pi(x, y) d\pi(x, y), \text{ if } \pi \ll dx \otimes dy \text{ and the integral exists and is finite} \\ -\infty \text{ otherwise} \end{cases}$$

Let $\pi_{M,T}$ denote a solution of (2.8); a proof of its existence can be found in [18] and references therein.

Fixing the temperature at 1, our aim in a first place is to find a matrix M such that $\hat{\pi}$ and $\pi_{M,1}$ have the same location in the covariogram; in other words they have the same cross-covariance matrix: $\sigma_{\hat{\pi}} = \sigma_{\pi_{M,1}}$. The gradient of W is given by the envelope theorem: $\nabla_M W(\cdot, T) = \sigma_{\pi_{M,T}}$. This remark implies that M is the solution of the following variational problem

$$\min_{M \in M_{I,J}(\mathbb{R})} W(M,T) - \sigma_{\hat{\pi}} \cdot M \tag{2.9}$$

$W(\cdot, T)$ is a convex function as a supremum of affine functions in M and consequently the objective function in (2.9) is convex as well: this is a classic unconstrained convex minimization problem. Figure 2.4 shows the projection of $\pi_{M,1}$ for a large number of randomly generated

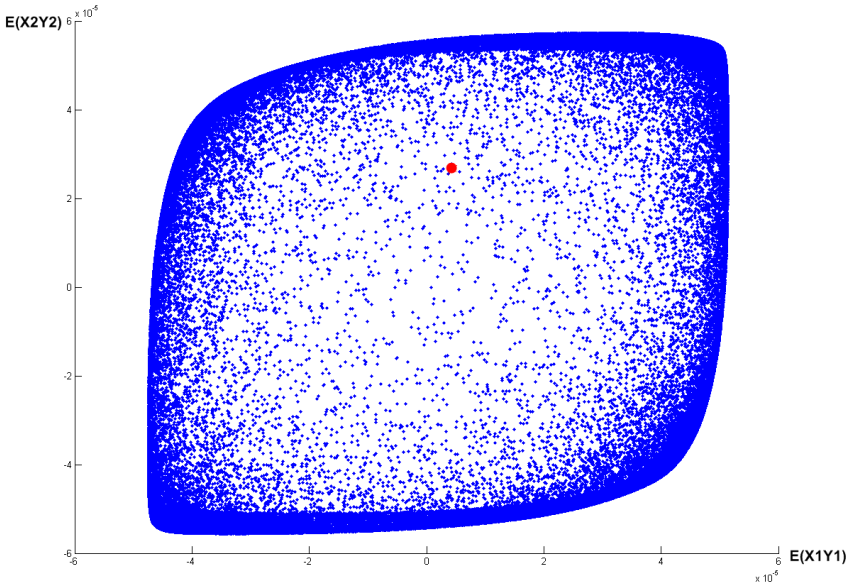

Figure 2.4: Projection of various π_M

matrices M. The bullet point is the projection of $\hat{\pi}$. One sees that any inner point of the covariogram can be attained by a properly chosen π_M.

2.4.2 Numerical solution

It can be shown that the optimal $\pi_{M,1}$ in (2.8) obeys a Schrödinger equation (see 2.9.3):

$$\log \pi_{M,1}(x,y) \quad = x'My + u(x) + v(y), \quad u \in L^1(dP), v \in L^1(dQ)$$

In other words, the log-likelihood of $\pi_{M,1}$ is the sum of a quadratic term $x'My$ and an additively separable function in x and y. The solution is found by setting u and v such that $\pi_{M,1}$ has

the marginals P and Q. This is the purpose of the well known (Deming & Stephan 1940, Von Neumann 1950) Iterative Projection Fitting Algorithm.

Let us recall in a few words the principle of it; we refer the interest reader to [18] for a more detailed exposition and a complete proof of the convergence. This algorithm consists in building a sequence π_n such that π_{2n} has first marginal P and π_{2n+1} has second marginal Q. It can be interpreted as Von Neuman's Iterated Projection algorithm with respect to the Kullback-Leibler distance. Its most remarkable property is the convergence of π_n towards a probability π with correct marginals P and Q. π_n has the following form:

$$\pi_{2n}(x,y) \propto e^{x'My+u_n(x)+v_n(y)} \quad \text{while} \quad \pi_{2n+1}(x,y) \propto e^{x'My+u_{n+1}(x)+v_n(y)}$$

The algorithm proceeds as follow: first choose some starting (u_0, v_0) defining π_0; for instance $v_0 = -y^2$ and $u_0 = -x^2$. We then look for some joint distribution π_1 whose first marginal is P, taking the form $e^{x'My+u_1(x)+v_0(y)}$. This writes

$$e^{u_1(x)} = \frac{P(x)}{\int e^{x'My+v_0(y)}dy}$$

Then we want to set v_1 so that $\pi_2(x,y) = e^{x'My+u_1(x)+v_1(y)}$ has second marginal Q and we get:

$$e^{v_1(y)} = \frac{Q(y)}{\int e^{x'My+u_1(x)}dx}$$

and so on, the recursion at step n writes

$$\left\{ \begin{array}{rcl} e^{u_{n+1}(x)} & = & \frac{P(x)}{\int e^{x'My+v_n(y)}dy} \\ e^{v_{n+1}(y)} & = & \frac{Q(y)}{\int e^{x'My+u_{n+1}(x)}dx} \end{array} \right.$$

This algorithm is typically a *fixed-point algorithm*; it finds (u, v) such that

$$\left\{ \begin{array}{rcl} \int e^{x'My+u(x)+v(y)}dy & = & P(x) \\ \int e^{x'My+u(x)+v(y)}dx & = & Q(y) \end{array} \right.$$

This builds a series of (u_n, v_n) (defined up to a constant) which enjoys a convergence property: $\pi_n \to \pi$, in total variation (again we refer to [18] for more details). An important remark is that in the case of discrete distributions P and Q, the previous formulae become simply:

$$\left\{ \begin{array}{rcl} e^{v_{n+1}(y)} & = & \frac{Q(y)}{\sum_x r(x,y)e^{u_n(x)}} \\ e^{u_{n+1}(y)} & = & \frac{P(x)}{\sum_y r(x,y)e^{v_{n+1}(y)}} \end{array} \right.$$

where $r(x,y) = \frac{e^{x'My}}{\sum_{x,y} e^{x'My}}$. Eventually the convex minimization problem (2.9) can be solved by any gradient descent type algorithm. The BFGS algorithm is used in the examples below.

2.4.3 Derivation of the extreme coupling

We recall that our aim is to associate an inner coupling (i.e. a coupling which projects inside the covariogram) to some extreme coupling which projects onto the boundary of the covariogram, by finding a trajectory of couplings that goes smoothly from the inner one to the extremal one.

The previous algorithm yields a particular matrix $\hat{M}$ and a coupling $\pi_{\hat{M}}$ such that $\sigma_{\hat{\pi}} = \sigma_{\pi_{\hat{M},1}}$. This coupling was found by setting arbitrarily the temperature at 1; the entropy penalization was thus effective and this allowed to reach inner points in the covariogram. This temperature parameter is easily explained. When it goes to $+\infty$, the entropy penalization is predominant in (2.8). Informally, the solution coupling is the one exhibiting the more *disorder*: this is the independence coupling. On the contrary, the less is the temperature, the closer (2.8) is to the non penalized problem. Hence, the lower T, the more $\pi_{\hat{M},T}$ projects near the boundary. Hence associating $\hat{\pi}$ with an extreme coupling can be done in the following way: once $\hat{M}$ is found, a sequence of $\pi_{\hat{M},T_n}$, $T_n \downarrow 0$ yields on the covariogram a trajectory of points which tend to the boundary. Figure 2.5 summarizes this idea: each point on the curve is the projection

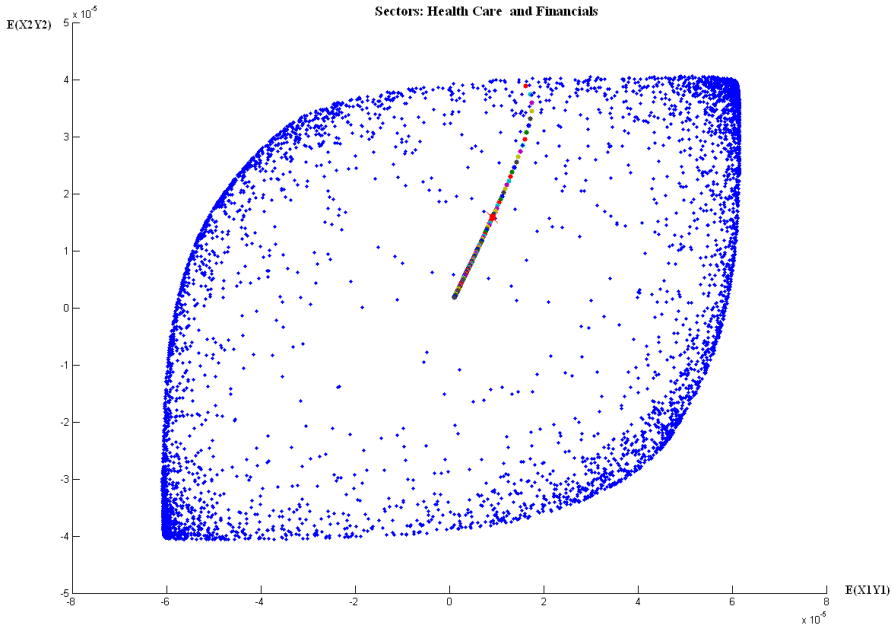

Figure 2.5: A trajectory toward an extreme coupling when the sectors are Health Care and Financials

of a $\pi_{\hat{M},T_n}$. As $T \to +\infty$, we recover the independence coupling whose projection is located at (0,0). When the temperature decreases, the trajectory passes on $\hat{\pi}$ at $T = 1$, and gradually approaches the boundary of the covariogram. Thus, the temperature can be seen as a means to control the strength of the dependence. This can be used to define formally an index of

dependence : choosing a norm $||\cdot||$ over the set of matrices $M_{I,J}(\mathbf{R})$ and using the homogeneity of W, namely $W(\lambda M, \lambda T) = \lambda W(M,T)$ for all $\lambda \in \mathbf{R}$, we have $\pi_{\hat{M},1} = \pi_{\hat{M}/||\hat{M}||,||\hat{M}||}$ and the temperature $1/||\hat{M}||$ appears as an indicator of the strength of the dependence between the marginals of $\hat{\pi}$. The matrix $\hat{M}$ can be seen as an affinity matrix : in the limit of $T \to 0$, the extreme coupling $\pi_{\hat{M},0}$ achieves the supremum of $\mathbf{E}_\pi(X'\hat{M}Y)$. $\hat{M}$ is thus the linear transform that makes X the most dependent with $\hat{M}Y$ under $\pi_{\hat{M},0}$.

2.5 Applications

In the financial applications below, the previous technique is applied to times series of linear daily returns on sectors of mainstream indices: S&P 500 and DJ Eurostoxx. We consider Health Care, Financial and Food & Beverage sectors of these indices: P and Q are distributions on $\mathbb{R}^3$. The historical data spans 5 years between September 2004 and September 2009. Table 2.1 gives summary statistics (the three first variables corresponds to S&P sectors, the last third

Table 2.1: Summary Statistics

Mean Returns	10^{-4} (1.03 −1.13 1.67)
Variance	10^{-4}. (1.36 7.65 1.16)
Correlation matrix	$\begin{pmatrix} 1 & & \\ 0.66 & 1 & \\ 0.76 & 0.62 & 1 \end{pmatrix}$
Cross-Covariance	10^{-5}. $\begin{pmatrix} 2.74 & 3.05 & 2.13 \\ 6.04 & 1.8 & 5.52 \\ 2.66 & 4.62 & 2.56 \end{pmatrix}$

to Eurostoxx). In particular, the correlations between sectors belonging to different indices are mild ($< 35\%$ in every case). Inside an index, correlation is well higher, but remains below 80%; this motivated our choice for these sectors: the marginal laws are not degenerated.

2.5.1 Numerical Results

P and Q are discrete distributions with equally weighted atoms in $\mathbb{R}^3$, each atoms being a vector of the returns at some date of the three sectors. The atoms are equally weighted as we consider that the daily returns are i.i.d random variables.

$$P = \frac{1}{N}\sum_{t=1}^{N} \delta_{r_t^X} \quad , \quad r_t^X = \text{vector of the linear returns on the S\&P500}$$

The optimal $\hat{M}$ we find when considering all three sectors or only Construction and Health Care are:

# of components	2	3
optimal M	$\begin{pmatrix} 0.23 & -0.14 \\ -0.10 & 0.40 \end{pmatrix}$	$\begin{pmatrix} 0.25 & -0.139 & -0.37 \\ -0.39 & 0.44 & -0.80 \\ -0.57 & -0.15 & 0.86 \end{pmatrix}$
error $= \frac{\|\sigma_M - \sigma_{\hat{\pi}}\|}{\|\sigma_{\hat{\pi}}\|}$	$\approx 0.1\%$	$< 0.2\%$

The linear returns are expressed in percentage. The error is computed as the percentage of difference between $\sigma_{\hat{\pi}}$, the cross-covariance target, and $\sigma_{\pi_{M,1}}$, the covariance matrix of the optimal coupling. They should be perfectly equal in theory and this percentage measures the convergence of the gradient algorithm.

2.5.2 Financial applications

The first application exploits further the affinity matrix $\hat{M}$. It consists in performing a singular value decomposition on it, in order to deduce indices of maximal correlation; it is related to the notion of canonical correlation.

The second one is based on considering the trajectory of couplings $T \mapsto \pi_{\hat{M},T}$ as a continuous family of scenarios of increasing dependence. They are used to build scenarios of stress-tests involving multivariate variables that can be useful for risk management. By stress-testing, we mean increase the index of dependence defined above (that is, lowering the temperature parameter), thus shifting away continuously from some coupling $\hat{\pi}$ to the extreme coupling $\pi_{\hat{M},0}$. This is to be compared to the method that consists in picking the maximum correlation coupling as the 'highest dependence scenario'; indeed this coupling might be less in line with the cross-covariance structure of the empirical coupling $\hat{\pi}$, yielding unexpected and undesired results when managing risky portfolios or options on several assets. Typically, we expect the cross-covariance matrix of $\pi_{\hat{M},0}$ to be located nearer from $\sigma_{\hat{\pi}}$ than the projection of the maximum correlation coupling.

Indices of maximal correlation

Recall that canonical correlation analysis consists, for two random vectors X and Y, in finding vectors a and b such that $(a'X, b'Y)$ solves $\max_{a,b} corr(a'X, b'Y)$. The *first canonical correlation*, defined as this maximum, is the highest diagonal element of the diagonal matrix that appears in the singular value decomposition of the matrix $\sigma_{XX}^{-1/2}\sigma_{XY}\sigma_{YY}^{-1/2}$ (see [13]). Let $\hat{M}$ be the affinity matrix of the coupling (X, Y). The singular value decomposition of this matrix writes $\hat{M} = USV'$, with U and V two orthogonal matrices and S a diagonal matrix with nonnegative entries. In particular,

$$\mathbf{E}_{\pi_{\hat{M},0}}\Big((\sqrt{S}U'X)'(\sqrt{S}V'Y)\Big) = \max_{\pi \in \Pi(P,Q)} \mathbf{E}_{\pi}\Big((\sqrt{S}U'X)'(\sqrt{S}V'Y)\Big)$$

In other words, if $(\tilde{X}, \tilde{Y}) = (\sqrt{S}U'X, \sqrt{S}V'Y)$, then this linear transform of (X, Y) has maximum covariance (under the law $\pi_{\hat{M},0}$). Thus $\sqrt{S}U'$ and $\sqrt{S}V'$ are the analogue of the optimal a and b in the canonical correlation framework.

This transform is useful to understand the link between the extreme coupling $\pi_{\hat{M},0}$ and the maximum correlation coupling, the one that corresponds to $M = Id$ in (2.5). Indeed, if $\tilde{P}$ is the law of $\sqrt{S}U'X$ with $X \sim P$, $\tilde{Q}$ is defined likewise from Q, and $\tilde{\pi}_{\hat{M},0}$ is the law of $(\sqrt{S}U'X, \sqrt{S}V'Y)$ where $(X, Y) \sim \pi_{\hat{M},0}$, then $\mathbf{E}_{\tilde{\pi}_{\hat{M},0}}(X'Y) = \max_{\pi \in \Pi(\tilde{P},\tilde{Q})} \mathbf{E}_{\pi}(X'Y)$. The singular value decomposition of the affinity matrix provides linear transform of the marginals that makes the extreme coupling $\pi_{\hat{M},0}$ the maximum correlation coupling after a scaling of the marginals by these transforms.
As an example, in the case of the 3 components chosen above, this transform writes

$$\begin{aligned} \tilde{X} &= \begin{pmatrix} -0.42\ X_1 + 0.95\ X_2 - 0.019\ X_3 \\ -0.64\ X_1 - 0.27\ X_2 + 0.26\ X_3 \\ 0.11\ X_1 + 0.06\ X_2 + 0.35\ X_3 \end{pmatrix} \\ \tilde{Y} &= \begin{pmatrix} -0.30\ Y_1 + 0.99\ Y_2 - 0.13\ Y_3 \\ -0.67\ Y_1 - 0.16\ Y_2 + 0..28\ Y_3 \\ 0.12\ Y_1 + 0.08\ Y_2 + 0.34\ Y_3 \end{pmatrix} \end{aligned}$$

This result states that $\tilde{X}$ and $\tilde{Y}$ are most correlated to one another under the law of the extreme coupling. These two vectors are composed of portfolios involving the components of the original index and can be viewed as new indices: we speak of *indices of maximal correlation*. When the strength of dependence is maximal ($T = 0$), they maximize the correlation $\mathbf{E}(\tilde{X}\tilde{Y})$ among all others law of probability with same marginals.

Portfolios stress-testing

In order to underline the necessity of accounting properly for the multivariate dependence, the problem of one-period allocation is considered. Suppose a universe of allocation consists in a set of assets; the problem is to study the impact of the change of the dependence between two subsets of this universe. They shall be denoted $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_m)$. In the examples below, the assets are S&P Sector Indices, and X is composed of Materials, Construction and Retail indices, while Y is composed of Food and Beverage, Health Care, Financials and Utilities indices. The corresponding summary statistics are given in table 2.2. Correlation is higher than in the above examples as the sectors are industrial sectors on a single index, the S&P500.

Consider an investor solving a classic Markowitz allocation problem, with an investment horizon of one year: $\max_{\sum_i w_i = 1} \mu \cdot w - \frac{\lambda}{2} w' \Sigma w$. μ are the expected yearly returns of the stocks and Σ the covariance matrix of the returns. We assume that both μ and Σ are the standard empirical estimators (in other words, the investor do not make any guess as to the future behavior of the assets), computed over a period of one-year, the in-sample period. The risk

Table 2.2: Summary Statistics

Mean Returns	$10^{-4}.\begin{pmatrix} 2.89 & 1.67 & 1.03 & -1.13 & 1.97 & 2.01 & 1.85 \end{pmatrix}$
Variance	$10^{-4}.\begin{pmatrix} 3.59 & 1.16 & 1.36 & 7.65 & 1.92 & 0.984 & 3.25 \end{pmatrix}$
Correlation matrix	$\begin{pmatrix} 1 & & & & & & \\ 0.72 & 1 & & & & & \\ 0.71 & 0.76 & 1 & & & & \\ 0.69 & 0.86 & 0.65 & 1 & & & \\ 0.69 & 0.85 & 0.69 & 0.76 & 1 & & \\ 0.69 & 0.67 & 0.75 & 0.62 & 0.66 & 1 & \\ 0.70 & 0.76 & 0.60 & 0.72 & 0.74 & 0.56 & 1 \end{pmatrix}$
Cross-Covariance	$10^{-4}.\begin{pmatrix} 1.41 & 1.53 & 3.62 & 1.85 \\ 0.921 & 0.979 & 1.83 & 1.05 \\ 1.27 & 1.45 & 3.73 & 1.50 \end{pmatrix}$

aversion parameter λ is set at 3. The solution to the Markowitz allocation problem with these parameters is denoted w. The risk of a portfolio is here identified to its variance, and is known as soon as the covariance between the assets is specified. When performing the allocation at time 0, the investor is expecting a risk of $w'\Sigma w$. The stress-test consists in considering that the market conditions changes after the investment decision: the strength of dependence between X and Y increases.

The affinity matrix is computed with respect to the in-sample data. The whole trajectory of couplings toward the boundary obtains, parameterized by the temperature T. These couplings π_T yield stressed covariance matrices $\Sigma_T = \mathbf{E}_{\pi_T}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))')$. Σ_T represents a scenario where the marginals of X and Y are left unchanged, while the realized dependence between X and Y has increased, compared to the initial covariance matrix Σ. In a first place, the expected risk of the portfolio, $w'\Sigma w$, is compared to the realized yearly risk $w'\Sigma_T w$. It gives a first hint as to unexpected risks the investor might face when the dependence varies and the allocation decision does not forecast this change. The graph 2.6 shows this effect. The variance obtained at temperature 1 is $w'\Sigma w$; in the worst case, where the realized covariance is $\Sigma_{0.1}$, the investor chooses a portfolio that yields an extra 4% of variance than expected. When the dependence is properly accounted for, the investor determines the optimal weights w_T according to the covariance Σ_T. The opportunity cost $\mu \cdot w_T - \mu \cdot w$ is the loss in term of returns that arises when the dependence increases, while the investor sticks to the allocation w. This cost is more and more significant as the temperature lowers, reaching 6% in this case. A comparison with the usual extreme multivariate coupling, namely the maximum correlation coupling is enlightening. First of all, this coupling is not defined when the dimension of X and Y are different. Consequently an asset is removed from Y and the same computation as above is performed: a covariance matrix Σ_B that would be the realized covariance if the assets were in maximum correlation dependence is computed. On this particular example, the variance $w'\Sigma_B w$ is 60% *lower* than the expected variance $w'\Sigma w$. Other examples can yield to a significantly higher covariance. This shows that the maximum correlation coupling might not be always adapted as a means of stress-testing the dependence. A more classical way to stress the dependence is to suppose that the correlation between X_i and Y_j is ρ for all i and

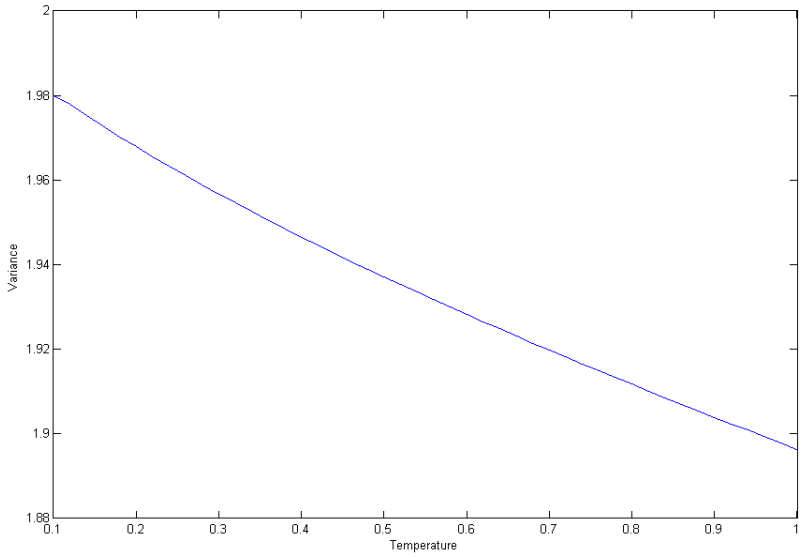

Figure 2.6: Plot of $T \mapsto w'\Sigma_T w$

j; the correlation matrix between X and Y is filled with ρ and the resulting cross-covariance matrix is denoted Σ_ρ. A first problem of this method is that it is known beforehand that, depending on the marginals, Σ_ρ might not be an admissible cross-covariance matrix for P and Q; the resulting variance-covariance matrix of the vector (X, Y) might fail to be semi-definite positive. This stress-test yields in this case underestimated risks. Indeed, while in our framework the variance $w'\Sigma w$ is at 1.91, this level of variance is attained only when ρ is above 95%, while the mean of the empirical cross-correlation is around 60%. Furthermore, even if ρ is set at 100% (disregarding the admissibility problem evoked above), the resulting variance is still lower than the one obtained with the extreme coupling.

It appears that the trajectory $T \mapsto \pi_T$ provides a coherent sequence of covariance matrices Σ_T that models a rise in the dependence between X and Y. This method respects both marginals and has the advantage of generating admissible matrices where the usual method of parameterizing correlation matrices by a single parameter could yield incoherent covariance matrices. Moreover, the maximum correlation coupling fails in this setting to properly account for increasing risk of dependence, likely because it ignores the cross-correlation effects.

Options pricing

This method of increasing the multivariate dependence can be also applied for rainbow options (options on several underlyings) pricing. As a case study, consider the underlyings $X_1, \ldots, X_n$,

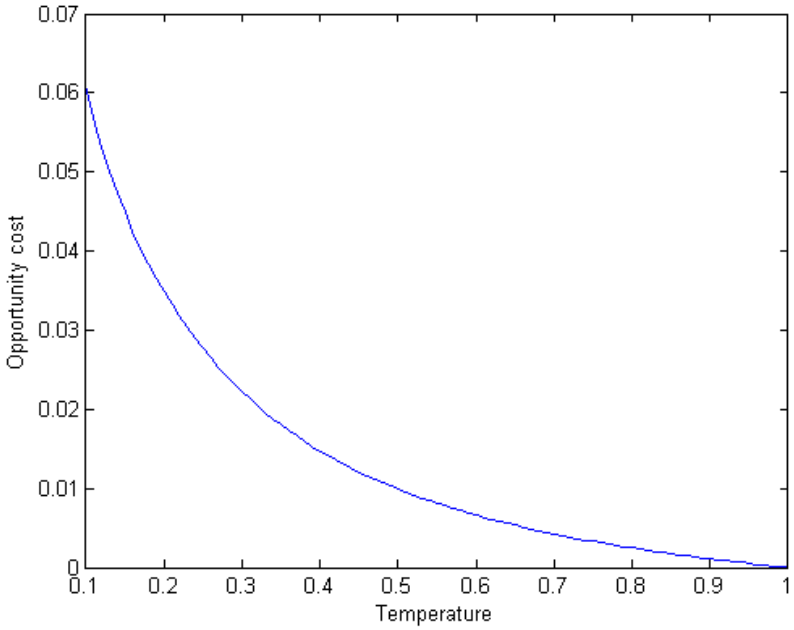

Figure 2.7: Opportunity cost as a function of the temperature

$Y_1, \ldots, Y_m$. It is assumed that they all follows log-normal diffusions, with parameters

$$\begin{cases} \frac{dX_t^i}{X_t^i} & = \mu_i^X dt + \sigma_i^X dW_t^i \quad , \quad d\langle W^i, W^j \rangle_t = \rho_{ij}^X dt \\ \frac{dY_t^i}{Y_t^i} & = \mu_i^Y dt + \sigma_i^Y dB_t^i \quad , \quad d\langle B^i, B^j \rangle_t = \rho_{ij}^Y dt \end{cases}$$

The models is fully specified as soon as the correlation matrix between W and B is set. Consider the option that pays $\min((\max_i X_T^i - K)_+, (\max_j Y_T^j - K)_+)$; it is the minimum between the payoffs of two best-of options on the X^i on the one hand and the Y^j on the other hand. It pays when the X_T^i and Y_T^i perform well, but mitigates the gain by selecting the lowest payoff between $(\max_i X_T^i - K)_+$ and $(\max_j Y_T^j - K)_+$. The terminal distribution of the underlyings is discretized; the discrete marginals of vectors X and Y obtains. Their atoms are respectively denoted x_T^i and y_T^j. For each specification of a cross-covariance matrix A between X and Y, a trajectory $\pi_T(A)$ is obtained as well as a series of prices:

$$\begin{aligned} P_T(A) &= \mathbf{E}_{\pi_T(A)}\Big(\min((\max_i X_T^i - K)_+, (\max_j Y_T^j - K)_+) \Big) \\ &= \sum_{i,j} \min((\max_i x_T^i - K)_+, (\max_j y_T^j - K)_+) \pi_T(A)(x_T^i, y_T^j) \end{aligned}$$

In the following example, X has 3 components and Y has 4. The riskless rate is constant and set at zero; μ_X and μ_Y are supposed to have null drift (i.e. we suppose that the above dynamics is given with respect to the risk-neutral measure), $\sigma^X = (0.15, 0.20, 0.22)'$ and

$\sigma^Y = (0.13, 0.10, 0.16, 0.18)'$. The correlation structure is set as follows; for the sake of the exposition W and B are standard Brownian motions ($\rho^X = Id_n$ and $\rho^Y = Id_m$) while the cross-correlation matrix between W and B is randomly generated, and set at

$$\begin{pmatrix} 0.087 & 0.126 & 0.068 & 0.100 \\ 0.490 & 0.438 & 0.006 & 0.149 \\ 0.136 & 0.369 & 0.447 & 0.331 \end{pmatrix}$$

The strike is set at 1, i.e. at time 0 the option is at-the-money.

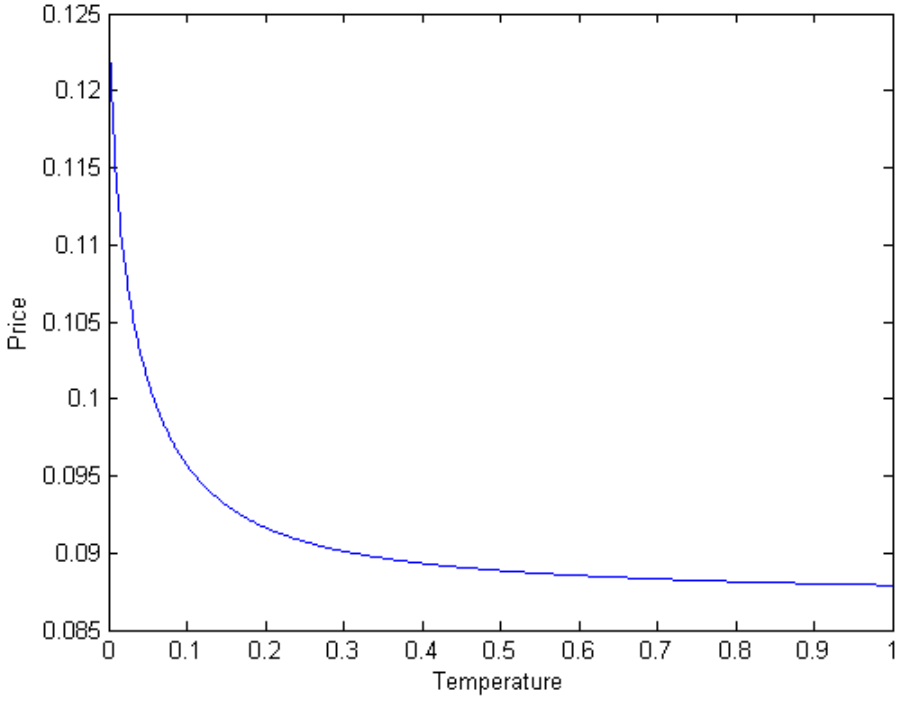

Figure 2.8: Price as a function of the temperature

As seen on graph 2.8, the price increases as the temperature lowers; this is an expected behavior, as when the dependence between the assets increases, so does the dependence between their respective maxima and hence the minimum of these maxima tends to be higher, which yields a higher price. In this setting, the stress-test increases the price by more than 30%. This must be compared to the price that is obtained when the cross-correlation matrix is taken of the form $\Sigma_\rho = \begin{pmatrix} \rho & \dots & \rho \\ \vdots & & \vdots \\ \rho & \dots & \rho \end{pmatrix}$. As a matter of fact, the stress-test of the cross-correlation fails, as the resulting correlation matrix $\begin{pmatrix} Id & \Sigma_\rho \\ \Sigma_\rho & Id \end{pmatrix}$ is no longer definite positive when $\rho > \frac{1}{2\sqrt{3}}$ which is lower than 30%. And even in the limit $\rho \to \frac{1}{2\sqrt{3}}$, the price does not reach 0.075, and is still lower than the non-stressed price.

2.6 Conclusion

A recurring complaint in Applied Statistics is the "curse of dimensionality": models that have a simple, computationally tractable form in dimension one become very complex, both computationally and conceptually in higher dimension. We show here that convex analysis, along with the theory of Optimal Transport, can lead to efficient solutions to problem of extreme dependence. Building on a natural geometric definition of extreme dependence, we have introduced an index of dependence and used the latter to build stress-tests of dependence between two sets of economic variables. This is particularly relevant in the case of international finance, where the dependence between many economic variables in two countries is of interest.

Acknowledgments

The authors thank Rama Cont for a question which was the starting point of this chapter and Guillaume Carlier and Alexander Sokol for helpful conversation.

Bibliography

[1] K. Arrow. Uncertainty and the welfare of medical care. *Amer. Econom. Rev.*, 53:941–973, 1963.

[2] K. Arrow. *Essays in the Theory of Risk-Bearing*. North-Holland Publishing Co., Amsterdam, 1970.

[3] K. Borch. Equilibrium in a reinsurance market. *Econometrica*, 30:424–444, 1962.

[4] G. Carlier, R.-A. Dana, and A. Galichon. Pareto efficiency for the concave order and multivariate comonotonicity. *Journal of Economic Theory*, 147:207–229, 2012.

[5] G. Dall'Aglio. Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Sc. Norm. Super. Pisa*, 10:35–74, 1956.

[6] W. Deming and F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11:427–444, 1940.

[7] I. Ekeland, A. Galichon, and M. Henry. Comonotonic measures of multivariate risks. *Mathematical Finance*, 22:109–132, 2012.

[8] K. Fan. Fixed-point and minimax theorems in locally convex topological linear spaces. *Proceeding of the National Academy of Sciences*, 38, 1951.

[9] A. Galichon and M.Henry. Dual theory of choice under multivariate risks. *Journal of Economic Theory*, Forthcoming.

[10] A. Galichon and B. Salanié. Matching with trade-offs: Revealed preferences over competing characteristics. Working paper, 2010.

[11] J. J. Kosowsky and A. L. Yuille. The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Networks*, 7:477–490, 1994.

[12] M. Landsberger and I. Meilijson. Co-monotone allocations, Bickel-Lehmann dispersion and the Arrow-Pratt measure of risk aversion. *Ann. Oper. Res.*, 52:97–106, 1994.

[13] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, 1979.

[14] R. B. Nelsen. *An Introduction to Copulas*. Springer, 2006.

[15] G. Puccetti and M. Scarsini. Multivariate comonotonicity. *Journal of Multivariate Analysis*, 101:291–304, 2010.

[16] S. Rachev. *Probability Metrics and the Stability of Stochastic Models*. John Wiley & Sons Ltd., 1991.

[17] L. Rüschendorf. Fréchet-bounds and their applications. In *Advances in Probability Distributions with Given Marginals*, volume 67 of *Math. Appl.*, pages 151–187. Kluwer Acad. Publ., 1990.

[18] L. Rüschendorf. Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, 23:1160–1174, 1995.

[19] D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57:571–587, 1989.

[20] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.

[21] M. Yaari. The dual theory of choice under risk. *Econometrica*, 55:95–115, 1987.

[22] V. Zolotarev. Probability metrics. *Theory Probab. Appl.*, 28:278–302, 1983.

2.7 Appendix

2.8 Facts on conic orders

If $K \subset \mathbf{M}_{I,J}(\mathbb{R})$ is a closed convex cone, a *base* for K is a convex set C with $0 \notin \bar{C}$ (the closure of C) and K is generated by C, i.e. $K = \mathbb{R}_+ C$. Thereafter C is supposed *compact*.

The *dual cone* associated to K is

$$K^* = \{\Sigma \in \mathbf{M}_{I,J}(\mathbb{R}) | \Sigma \cdot M \geq 0,\ M \in K\}$$

Its interior is also of interest, and is simply

$$K_+^* := Int(K^*) = \{\Sigma \in \mathbf{M}_{I,J}(\mathbb{R}) | \Sigma \cdot M > 0,\ M \in K \backslash \{0\}\}$$

Note that in both definitions, one can replace K and $K\backslash\{0\}$ with C.

A *strict partial order* is defined on E by setting

$$A \succ_K B \overset{\text{def}}{\Leftrightarrow} A - B \in K_+^*$$

If S is a subset of $\mathbf{M}_{I,J}(\mathbb{R})$, a *maximal* element of S for this order is $A \in S$ such that for all $B \in S$, $A - B \notin K_+^*$: A can not be 'strictly dominated' by any element in S.
These definitions apply of course when $\mathbf{M}_{I,J}(\mathbb{R})$ is replaced by any euclidean space.

2.9 Proof of the results

2.9.1 Proof of Theorem 3

Proof: As the covariogram is a closed convex set, a point $x \in \mathbf{M}_{I,J}(\mathbb{R})$ lies on its boundary if and only if there exists a nonzero $M \in \mathbf{M}_{I,J}(\mathbb{R})\backslash\{0\}$ such that $M \cdot x$ is maximal as a function of x. This translates the fact that there exists a supporting hyperplane at x. Thus σ_π is on the boundary of the covariogram iff there exists $M \in \mathbf{M}_{I,J}(\mathbb{R})\backslash\{0\}$ such that

$$M \cdot \sigma_\pi = \sup_{\tilde{\pi} \in \Pi(P,Q)} M \cdot \sigma_{\tilde{\pi}}$$

(where it is recalled that $M \cdot \sigma_\pi = Tr\,(M' \sigma_\pi)$).

Equivalence between (ii) and (iii) follows from a well-known result in Optimal Transport theory, the Knott-Smith optimality criterion (see [20], Th. 2.12). □

2.9.2 Proof of Theorem 4

Before we give the proof of the theorem, we state and prove a number of auxiliary results which are of interest per se.

First, in the case of a generic compact base C, we have a crucial, although technical, variational characterization of the maximality of σ_π:

Proposition 3 (Variational characterization of maximality)

$$\sigma_\pi \textit{ maximal iff } \sup_{\tilde{\pi} \in \Pi(P,Q)} \inf_{M \in C} (\sigma_{\tilde{\pi}} - \sigma_\pi) \cdot M = 0$$

In other terms, a coupling is maximal whenever there exists $M \in C$ such that σ_π maximizes $\sigma_{\tilde{\pi}} \cdot M$.

Proof: [Proof of proposition 3] First, note that for every $\pi \in \Pi(P,Q)$, the function

$$f : (\tilde{\pi}, M) \in \Pi(P,Q) \times C \mapsto (\Sigma_{\tilde{\pi}} - \Sigma_\pi) \cdot M$$

exhibits a saddlepoint $(\bar{\pi}, \bar{S})$:

$$\max_{\tilde{\pi} \in \Pi(P,Q)} \min_{M \in C} f(\tilde{\pi}, M) = f(\bar{\pi}, \bar{M}) = \min_{M \in C} \max_{\tilde{\pi} \in \Pi(P,Q)} f(\tilde{\pi}, M) \tag{2.10}$$

This is a consequence of a classical minmax theorem by Fan [8]: a continuous function over a product of compacts convex sets embedded in normed linear spaces, which is linear in both arguments exhibits a saddlepoint. Both $\Pi(P,Q)$ and C are compacts and convex. The compacity is C is an hypothesis and a well-known fact for $\Pi(P,Q)$, see [20] for instance. Moreover f is linear in M and $\tilde{\pi}$, and continuous in both arguments. Finally, $\Pi(P,Q)$ can be embedded in the space of Radon measures over $\mathbb{R}^I \times \mathbb{R}^J$ endowed with the bounded Lipschitz norm. We refer to Villani [20] chapter 7. for more details on this: the important thing is that $\Pi(P,Q)$ is a compact subset (for the norm) within this space.

Back to the proof of the result. If σ_π is maximal, then for all $\sigma_{\tilde{\pi}}$ one has $\sigma_{\tilde{\pi}} - \sigma_\pi \notin K_+^*$, which means that for some $M \in C$, $(\sigma_{\tilde{\pi}} - \sigma_\pi) \cdot M \leq 0$, hence

$$\sup_{\Pi(P,Q)} \inf_{C} (\sigma_{\tilde{\pi}} - \sigma_\pi) \cdot M \leq 0$$

Thanks to the compacity of K, we can apply the minmax formula 2.10 to invert the supremum and the infimum, and conclude the proof of one implication. On the contrary, if σ_π is not maximal then there exists some coupling $\tilde{\pi}$ such that $\sigma_{\tilde{\pi}} - \sigma_\pi \in K_+^*$. Thanks again to the compacity of C, $\inf_C (\sigma_{\tilde{\pi}} - \sigma_\pi) \cdot M > 0$ and the reverse implication is proved. $\square$

As a consequence, we are now ready to prove theorem 4.

Proof: [Proof of theorem 4] Because of the previous proposition, a coupling π such as

(X, MY) is an optimal transport plan shall satisfy

$$\mathbb{E}_\pi(X \cdot MY) = \sup_{\tilde{\pi} \in \Pi(P,Q)} \mathbb{E}_{\tilde{\pi}}(X \cdot MY)$$

As $\mathbb{E}_\pi(X \cdot MY) = \sigma_\pi \cdot M$, we conclude with the proposition 3.□

2.9.3 Schrödinger equation

An informal justification of the form of the solution to the entropic maximization problem is as follows. We assume that every coupling in $\Pi(P,Q)$ admits a density with respect to the Lebesgue measure on $\mathbb{R}^I \times \mathbb{R}^J$.

$$\max_{\pi \in \Pi(P,Q)} \mathbf{E}_\pi(X'MY) + TEnt(\pi) = \max_{\pi \in \Pi(P,Q)} \int x'My\pi(x,y) - T\log\pi(x,y)dxdy$$
$$= \max_{\pi \in \mathcal{M}_+^{\mathcal{E}}(\mathbb{R}^n \times \mathbb{R}^n)} \left\{ \min_{\substack{\phi \in L^1(dp) \\ \psi \in L^1(dq)}} \int (x'My - T\log\pi(x,y))\pi(x,y)dxdy \right.$$
$$\left. - \Big[\int (\phi(x) + \psi(y))d\pi(x,y) - \int \phi dp - \int \psi dq \Big] \right\} \quad \circledast$$

where $\mathcal{M}_+^{\mathcal{E}}(\mathbb{R}^I \times \mathbb{R}^J)$ is the set of nonnegative Radon measures on $\mathbb{R}^I \times \mathbb{R}^J$ for which the entropy is well-defined. Now the assumption on the marginals is relaxed, a sloppy way to get the result is to say that the solution should satisfy

$$\frac{\partial}{\partial \pi(x,y)} \min_{\substack{\phi \in L^1(dp) \\ \psi \in L^1(dq)}} \int [x'My - T\log\pi(x,y) - (\phi(x) + \psi(y))]\pi(x,y)dxdy = 0$$

If we could apply the envelope theorem, we would have the existence of a couple (ϕ^*, ψ^*) such that

$$x'My - T(1 + \log\pi(x,y)) - \phi^* - \psi^* = 0$$

which yields the expected form for π.

Here is a rigorous proof in the case where P and Q are absolutely continuous with respect to the Lebesgue measure.
The problem (2.8) is equivalent to solve the following minimization problem:

$$\min_{\Pi(P,Q)} \int \log\left(\frac{\pi(x,y)}{e^{x'My - |x|^2 - |y|^2} / \int e^{x'My - |x|^2 - |y|^2} dxdy} \right) \pi(x,y)dxdy$$

The quantity inside the min is the Kullback-Leibler distance (or relative entropy) of the distribution μ with density proportional to $e^{x'My - |x|^2 - |y|^2}$ (the $-|x|^2 - |y|^2$ ensures the integrability)

with respect to π. Minimizing this distance consists in projecting μ onto $\Pi(P,Q)$ with respect to the Kullback-Leibler distance. This is the purpose of IPFP. Rüschendorf [18] applies and states that the unique solution to this problem is of the form:

$$\pi^*(x,y) = a(x)b(y)e^{x'My-|x|^2-|y|^2}$$

with a and b two positive functions, which is the desired result.□

Chapter 3

Coupling Markovian diffusions with copulas

3.1 Introduction

Copulas are functions that represent the dependence of multivariate laws of probability. Namely, if $X_1, \dots, X_n$, $n \geq 2$, are real random variables on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$, their cumulative distribution function (cdf) is defined, for $(x_1, \dots, x_n) \in \mathbf{R}^n$, by $F(x_1, \dots, x_n) = \mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$. The copula approach to dependence consists in scaling the marginals X_i by their respective cdfs F_i; the cdf of the scaled vector is the copula and is defined, for $(u_1, \dots, u_n) \in [0,1]^n$, by $C(u_1, \dots, u_n) = \mathbf{P}(F_1(X_1) \leq u_1, \dots, F_n(X_n) \leq u_n)$. As $F_i(X_i)$ follows the uniform law on $[0,1]$, C is the cdf of a vector of uniform random variables on $[0,1]$. Eventually, the initial cdf can be written

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \tag{3.1}$$

and the copula C is uniquely determined on $ImF_1 \times \cdots \times ImF_n$. This fundamental statement is Sklar's theorem (c.f. the classic introductory book on copula by R. Nelsen [15]).

Copulas have been widely used, first in statistics where the notion was developed by Fréchet, Hoeffding and many others (see Nelsen's book [15] or Joe's book [10] and the numerous references therein), and then imposed itself as a convenient tool to model multivariate dependence in many fields. There has been a spectacular inflation of the use of copulas in financial mathematics in the last decade. This has been exemplified by the Gaussian copula model for the valuation of Credit Default Obligation by Li [13], and by numerous articles on valuation of derivatives on several underlyings, see for instance Cherubini and coauthors [4] and [3]. Since then, some drawbacks of the copula approach to dependence have been highlighted, see Mikosch [14]. An important criticism of copulas is their static nature, meaning that, whereas

they are suitable to describe the dependence between random variables, their use is more disputable when dependence modeling is needed at several dates, not to speak of continuous-time dependence modeling. Therefore, whereas they are useful to valuate financial products whose price depends only on the distribution of a vector of assets at a single time (such as European call options on several underlyings), they might not be the adequate tool when the dependence of the price with respect to the distribution of the assets is more complex. And indeed copulas have been applied less intensively in derivatives pricing than in risk management, where the static framework is more natural.

Nevertheless, a significant use of copulas in continuous-time setting was achieved by Darsow et al. [5]. For a real stochastic process $(X_t)_{t\in T}$, they obtained a formulation of the Chapman-Kolmogorov equation as an equation on the bivariate copulas C_{st}, which describes the dependence of the vector (X_s, X_t), $s \leq t$. This is remarkable as it allows for the specification of a Markov process by the one dimensional marginals (the law of X_t for each $t \geq 0$) and all the bivariate copulas C_{st}. However, these results do not generalize easily to the multivariate setting and describing both the time-dependence (dependence of X_s and X_t for all $s \leq t$) and the spatial dependence (dependence of $(X_t^1, \ldots, X_t^n)$ for all t) of a multivariate process is a complex problem which has been addressed by Cherubini et al. [3] in a discrete time setting. In discrete time, there exists also a substantial literature on dynamic copula models based on times-series, see e.g. Patton [16] and van den Goorbergh et al. [22].

This chapter takes a different route. It tackles the problem of coupling a pair of Markovian diffusions X_t and Y_t, and controlling the space-dependence, namely the copulas of (X_t, Y_t). This is done by assuming that the Brownian motions driving the diffusions, B and W, satisfy $d\langle B, W\rangle_t = \rho(t, X_t, Y_t)dt$ where $\rho(t, X_t, Y_t)$ is a correlation which depends on the state of the marginals diffusions. This type of model can be related to 'local correlation models', described in Langnau [12] and Reghai [18], although in these models the emphasis is put on calibrating the function ρ in order to match observed prices of various options on several underlyings, in the spirit of Dupire's local volatility model [6]. A partial differential equation, which first appeared in Galichon's [7], that describes the evolution of the copula C_{X_t,Y_t} is derived. In the case where the marginal diffusions are Brownian motions, this PDE allows to find explicit form of the correlation function in order for the resulting coupled Brownian motions to have a stationary (and possibly non Gaussian) copula. Moreover, it can be used to prove that several well-known copulas are unsuitable to couple Brownian motions. Eventually, this technique is applied to the simulation of a constant proportion portfolio insurance strategy (CPPI). This example aims at assessing the impact of copulas in the trigger probability of a CPPI in a coupled Black-Scholes model, where the driving Brownian motions are coupled by various copulas.

3.2 Coupling SDE and coupling copula

3.2.1 Correlated Brownian motions

This section recalls how Brownian motions can be coupled with a stochastic correlation function, and introduce the coupling correlation function, before defining the coupling of markovians diffusions.

Correlating Brownian motions with deterministic correlation The construction is classical: if $\rho_t \in [-1,1]$ is a measurable function on $\mathbf{R}_+$ which is locally square integrable, and (B^1, Z) is a two dimensional standard Brownian motion, then

$$B_t^2 = \int_0^t \rho_s dB_s^1 + \int_0^t \sqrt{1-\rho_s^2} dZ_s \tag{3.2}$$

is a Brownian motion (with respect to its natural filtration), and $\langle B^1, B^2\rangle_t = \int_0^t \rho_s ds$. Indeed, it is a continuous process, its quadratic variation $\langle B^2\rangle_t = t$ and it is a continuous local martingale. Hence it is a Brownian motion. Moreover, $\langle B^1, B^2\rangle_t = \int_0^t \rho_s d\langle B^1\rangle_s = \int_0^t \rho_s ds$ $\square$.

Correlated BM with stochastic correlation The previous construction extends to the case where ρ_t is a progressively measurable process with respect to the (augmented) filtration generated by (B^1, Z), and ρ_t is locally square integrable. Defining

$$dB_t^2 = \rho_t dB_t^1 + \sqrt{1-\rho_t^2} dZ_t \tag{3.3}$$

i.e. $B_t^2 = \int_0^t \rho_s dB_s^1 + \int_0^t \sqrt{1-\rho_s^2} dZ_s$, then if equation (3.3) has a strong and non explosive solution, B^2 is a Brownian motion, and $d\langle B^1, B^2\rangle_t = \rho_t dt$, just as in the deterministic case.

Correlated BM with coupling correlation function Consider a deterministic function $\rho_t(x,y)$ that is bounded by 1, and measurable. We would like to define a bidimensional Markov process (B^1, B^2):

$$d\langle B^1, B^2\rangle_t = \rho(t, B_t^1, B_t^2) dt \tag{3.4}$$

Accordingly, consider the following equation

$$dB_t^2 = \rho_t(B_t^1, B_t^2) dB_t^1 + \sqrt{1-\rho_t^2(B_t^1, B_t^2)} dZ_t \tag{3.5}$$

Assume that this diffusion equation has a strong solution. Then B^2 is a Brownian motion and satisfy the equation (3.4). The solution (B^1, B^2) of this SDE is called a *coupled Brownian motion*, and the function $(t,x,y) \in \mathbf{R}_+ \times \mathbf{R}^2 \mapsto \rho_t(x,y)$ is called *the coupling correlation*.

Coupling Markovian diffusions More generally, we want to give a meaning to the *coupling equation*

$$\begin{cases} dX_t &= a^X(t,X_t)dt + \sigma^X(t,X_t)dW_t^X \\ dY_t &= a^Y(t,Y_t)dt + \sigma^Y(t,Y_t)dW_t^Y \\ d\langle W^X, W^Y\rangle_t &= \rho_t(X_t,Y_t)dt \\ (X_0,Y_0) &\sim \mu_0 \end{cases} \tag{3.6}$$

where ρ is bounded by 1 and measurable. This equation corresponds to the coupling of two Markovian diffusions X_t and Y_t with the coupling correlation $\rho_t(X_t,Y_t)$. This equation is formulated unambiguously as

$$\begin{cases} dX_t &= a^X(t,X_t)dt + \sigma^X(t,X_t)dW_t^X \\ dY_t &= a^Y(t,Y_t)dt + \sigma^Y(t,Y_t)(\rho_t(X_t,Y_t)dW_t^X + \sqrt{1-\rho_t^2(X_t,Y_t)}dZ_t) \\ (X_0,Y_0) &\sim \mu_0 \end{cases}$$

where Z_t is a Brownian motion independent of W^X. Provided that the above equation admits a strong solution (X_t,Y_t) (classical conditions that ensure it are recalled in appendix 17), the process defined by

$$dW_t^Y = \rho_t(X_t,Y_t)dW_t^X + \sqrt{1-\rho_t^2(X_t,Y_t)}dZ_t$$

is indeed a Brownian motion, and $d\langle W^X, W^Y\rangle_t = \rho_t(X_t,Y_t)dt$.

3.2.2 A partial differential equation on the copulas

The Kolmorogov forward equation of a diffusion whose law at time $t>0$ has density f_t is an evolution equation of f_t. The purpose of this section is to show how, after scaling the marginal diffusions by their cdfs, a Kolmogorov forward equation for the coupling equation (3.6) is obtained; this equation describes the evolution of the copula C_t of the bivariate diffusion. This 'copula PDE' makes a link between the coupling correlation ρ_t and the copula C_t, and is the core of this chapter.

This section provides results of existence and uniqueness relative the Kolmogorov forward equation of the diffusion (3.6). Although these results might seem classical to the reader familiar with diffusion theory (as described for instance in the comprehensive book of Stroock and Varadhan [21] or in the more recent book of Stroock [20]), the scaling of the marginals must be handled carefully in order to derive the copula PDE rigorously, and this is the scope of the following results.

Notations and Hypotheses

Let (X_t, Y_t) be a strong solution to the coupling equation (3.6). The marginals cdfs F_t^X, F_t^Y are the cumulative distribution functions of X_t and Y_t, and, likewise, the marginals densities (when they exist) are denoted f_t^X and f_t^Y. These are always supposed positive thereafter, and thus the copula of (X_t, Y_t) is defined unambiguously and is denoted C_t.
The drifts and volatilities a^X, a^Y, σ^X, σ^Y are functions defined on $\mathbf{R}_+ \times \mathbf{R}$ that take values in $\mathbf{R}$ and are always assumed measurable and locally bounded (a function $f(t,x)$ is locally bounded if for all compact $K \subset \mathbf{R}_+ \times \mathbf{R}$, $\sup_{(t,x)\in K} |f(t,x)| < +\infty$.)
The set of twice continuously differentiable functions with compact support on $(0,1)^2$ is denoted $\mathcal{C}_c^2((0,1)^2)$.
$\mathcal{C}^{1,k}(\mathbf{R} \times (0,1)^2)$, k integer, is the set of functions $u(t,x)$ that are continuously differentiable k times in the space variable, continuously differentiable in the time variable, and such that $\partial_t u(t,x)$ is also continuously differentiable k times in x.
For a measure μ on $(0,1)^2$ and a function $\varphi \in \mathcal{C}_c^2((0,1)^2)$, the bracket $\langle \mu, \varphi \rangle$ is defined as $\int_{(0,1)^2} \varphi(x) d\mu(x)$. For a locally integrable function f, $\langle f, \varphi \rangle$ is defined as $\langle f dx, \varphi \rangle$.

A reminder on the Kolmogorov forward equation

Hypotheses for a Kolmogorov forward equation to hold in a sufficiently general case for our purpose are:

Proposition 4 *Let $\mathcal{L}_t$ be the infinitesimal generator of a Markovian diffusion X_t in $\boldsymbol{R}^N$, $N \geq 1$, and let P_t be the law of the diffusion at time t. Assume the drift and the volatility of X_t are locally bounded. Then the following equation holds: for all $\varphi \in \mathcal{C}_c^2(\boldsymbol{R}^N)$,*

$$\langle P_t, \varphi \rangle = \langle P_0, \varphi \rangle + \int_0^t \langle \mathcal{L}_s \varphi, P_s \rangle ds$$

This can be written informally

$$\partial_t P_t = \mathcal{L}_t^* P_t$$

where $\mathcal{L}_t^*$ denote the adjoint operator of $\mathcal{L}_t$.

This is a standard result and is basically an application of Itō's lemma.

Proof of the copula PDE

The scaling of the marginals does not make sense at time 0 when the initial distribution μ_0 has a singular component, and it is assumed in the first place that it admits a density with respect to the Lebesgue measure.

The two next lemmas prove that the scaled marginals remain diffusion processes, and identify the infinitesimal generator of the bivariate scaled diffusion.

Lemma 4 *Assume that the marginal cdfs are in* $\mathcal{C}^{1,2}(\boldsymbol{R}_+ \times \boldsymbol{R})$. *Assume moreover that* σ^X *and* σ^Y *are continuously differentiable in the space variable. Then the scaled marginal processes* $\tilde{X}_t = F_t^X(X_t)$ *and* $\tilde{Y}_t = F_t^Y(Y_t)$ *are diffusion processes. The infinitesimal generator of* $\tilde{X}_t$ *is*

$$\mathcal{L}_t^{\tilde{X}} = \frac{1}{2}\partial_x \left((f_t^X \sigma_t^X)((F_t^X)^{-1}(x))\partial_x \right)$$

The infinitesimal generator of $(\tilde{X}_t, \tilde{Y}_t)$ *is* $\mathcal{L}^{scaled} = \mathcal{L}_t^{\tilde{X}} + \mathcal{L}_t^{\tilde{Y}} + \mathcal{L}_t^{XY}$ *where* $\mathcal{L}_t^{\tilde{X}\tilde{Y}}$ *is the operator* $\varphi \in \mathcal{C}_c^2((0,1)^2) \mapsto \tilde{\rho}_t \tilde{f}_t^X \tilde{\sigma}_t^X \tilde{f}_t^Y \tilde{\sigma}_t^Y \partial_{uv}^2 \varphi$, *and* $\tilde{g}(t,u,v)$ *stands for* $g(t,(F_t^X)^{-1}(u),(F_t^Y)^{-1}(v))$.

Proof: As F_t^X and F_t^Y are $\mathcal{C}^{1,2}(\mathbf{R}_+ \times \mathbf{R})$, the Itō's lemma applies to $F_t^X(X_t)$ and $F_t^Y(Y_t)$, and we derive the diffusion equations for the scaled variables $\tilde{X}_t = F_t^X(X_t)$ and $\tilde{Y}_t = F_t^Y(Y_t)$:

$$\begin{aligned} d\tilde{X}_t &= f_t^X(X_t)dX_t + \partial_t F_t^X(X_t)dt + \frac{1}{2}(f_t^X)'(X_t)d\langle X,X\rangle_t \\ &= \Big[f_t^X(X_t)a^X(t,X_t) + \partial_t F_t^X(X_t) + \frac{1}{2}(f_t^X)'(X_t)(\sigma^X)^2(t,X_t) \Big]dt + f_t^X(X_t)\sigma^X(t,X_t)dW_t^X \end{aligned}$$

where $f_t^X = \partial_x F_t^X$ is the pdf of X_t. Because the cdfs are strictly increasing, $X_t = (F_t^X)^{-1}(\tilde{X}_t)$, and $\tilde{X}_t$ is a diffusion with the following dynamics:

$$d\tilde{X}_t = \Big[\widetilde{f_t^X a^X(t,\cdot) + \partial_t F_t^X + \frac{1}{2}(f_t^X)'(\sigma^X(t,\cdot))^2} \Big](\tilde{X}_t)dt + \widetilde{f_t^X \sigma^X(t,\cdot)}(\tilde{X}_t)dW_t^X$$

Let $\mu(t,x)$ and $vol(t,x)$ be the drift and the volatility in this equation. The density of $\tilde{X}_t$ is constant as it follows the uniform law on $[0,1]$. Let $\varphi \in \mathcal{C}_c^2((0,1)^2)$ and $\Phi(x) = \int_0^x \varphi(s)ds$. Then Itō's lemma yields

$$\begin{aligned} \mathbf{E}(\Phi(\tilde{X}_t)) = \mathbf{E}(\Phi(\tilde{X}_0)) &+ \mathbf{E}\Big(\int_0^t \varphi(\tilde{X}_s)\mu(s,\tilde{X}_s) + \frac{1}{2}\varphi'(\tilde{X}_s)vol^2(s,\mu(s,\tilde{X}_s))ds \Big) \\ &+ \mathbf{E}\Big(\int_0^t \varphi(\tilde{X}_s)vol(s,\tilde{X}_s)dW_s^X \Big) \end{aligned}$$

This last expected value is zero as the integrand is locally bounded and adapted with respect to W^X, and the integral is a martingale. As the law of $\tilde{X}_t$ is the uniform law on $(0,1)$ for all t, taking the time derivative of this expression yields:

$$0 = \langle \mu(t,\cdot), \varphi(x) \rangle + \frac{1}{2}\langle vol^2(t,\cdot), \partial_x \varphi(x) \rangle$$

for all $\varphi \in \mathcal{C}_c^2((0,1))$. As the function vol^2 is differentiable in the space variable by hypothesis,

$\mu(t,x) = \frac{1}{2}\partial_x vol^2(t,x)$ for all $(t,x) \in \mathbf{R}_+ \times (0,1)$, and the infinitesimal generator of the scaled diffusion is:

$$\mathcal{L}_t^{\tilde{X}}\varphi(t,x) = \frac{1}{2}\partial_x(vol(t,x)^2\partial_x\varphi)$$

□

The point to derive a PDE on the copula C_t is then to perform a mere integration by part. In order to simplify the exposition, it is assumed that the diffusion (X_t, Y_t) has a smooth copula.

Proposition 5 *Assume that the copula C_t of the coupled diffusion at time t is in $\mathcal{C}^{1,2}(\boldsymbol{R}_+ \times (0,1)^2)$. Assume that the hypotheses of lemma 4 hold: the marginal diffusions have continuously differentiable densities f_t^X and f_t^Y, and the volatilities are continuously differentiable in the space variable. Then the copula family C_t satisfies the following weak PDE: for all $t > 0$, for all $\varphi \in \mathcal{C}_c^2((0,1)^2)$,*

$$\begin{aligned}\langle \partial_t C_t, \partial_{uv}^2\varphi\rangle = \Big\langle \frac{1}{2}\Big(\Big(\widetilde{f_t^X\sigma^X}(t,\cdot)\Big)^2\partial_{uu}^2 C_t + \Big(\widetilde{f_t^Y\sigma^Y}(t,\cdot)\Big)^2\partial_{vv}^2 C_t\Big) \\ + (\rho_t f_t^X(\cdot)\sigma^X\widetilde{(t,\cdot)f_t^Y}(\cdot)\sigma^Y(t,\cdot))\partial_{uv}^2 C_t, \partial_{uv}^2\varphi\Big\rangle\end{aligned} \tag{3.7}$$

Proof: Let $\varphi \in \mathcal{C}_c^2((0,1)^2)$. The Kolmogorov forward equation states that:

$$\langle \mathbf{P}_{\tilde{X}_t,\tilde{Y}_t}, \varphi\rangle = \langle \mathbf{P}_{\tilde{X}_0,\tilde{Y}_0}, \varphi\rangle + \int_0^t \langle \mathcal{L}_s^{scaled}(\varphi), \mathbf{P}_{\tilde{X}_s,\tilde{Y}_s}\rangle ds$$

Equivalently,

$$\langle \partial_{uv}^2 C_t, \varphi\rangle = \langle \partial_{uv}^2 C_0, \varphi\rangle + \int_0^t \langle \mathcal{L}_s^{scaled}(\varphi), \partial_{uv}^2 C_s\rangle ds$$

We then detail the integrations by parts. Let $vol_t^{\tilde{X}}(u) = \widetilde{\sigma_t^X f_t^X}(u)$.

$$\begin{aligned}\langle \mathcal{L}_t^{\tilde{X}}(\varphi), \partial_{uv}^2 C_t\rangle &= \frac{1}{2}\int_0^1\int_0^1 \partial_u(vol_t^{\tilde{X}}(u)^2\partial_u\varphi(u,v))\partial_{uv}^2 C_t(u,v)dudv \\ &= \frac{1}{2}\int_0^1 \partial_u C_t(u,1)\partial_u(vol_t^{\tilde{X}}(\cdot)^2)\partial_u\varphi(\cdot,\cdot))(u,1)du \\ &\quad - \frac{1}{2}\int_0^1 \partial_u C_t(u,0)\partial_u(vol_t^{\tilde{X}}(\cdot)^2\partial_u\varphi(\cdot,\cdot))(u,0)du \\ &\quad - \frac{1}{2}\int_0^1\int_0^1 \partial_u C_t(u,v)\partial_u(vol_t^{\tilde{X}}(u)\partial_{uv}^2\varphi(u,v))dudv\end{aligned}$$

As φ has a compact support the two first integrals are zero. Another integration by parts in the last integral yields

$$-\int_0^1\int_0^1 \partial_u C_t(u,v)\partial_u(vol_t^{\tilde{X}}(u)\partial^2_{uv}\varphi(u,v))dudv = \int_0^1\int_0^1 \partial^2_{uu}C_t(u,v)vol_t^{\tilde{X}}(u)^2\partial^2_{uv}\varphi(u,v)dudv$$
$$-\int_0^1 [\partial_u C_t(1,v)vol_t^{\tilde{X}}(1,v)^2\partial^2_{uv}\varphi(1,v) - \partial_u C_t(0,v)vol_t^{\tilde{X}}(0,v)^2\partial^2_{uv}\varphi(0,v)]dv$$

Once again the boundary terms are zero, and eventually, we have

$$\langle \mathcal{L}_t^{\tilde{X}}(\varphi), \partial^2_{uv}C_t\rangle = \langle \frac{1}{2}(vol_t^{\tilde{X}})^2\partial^2_{uu}C_t, \partial^2_{uv}\varphi\rangle$$

Thus, for all $\varphi \in \mathcal{C}_c^2((0,1)^2)$,

$$\langle C_t, \partial^2_{uv}\varphi\rangle = \langle C_0, \partial^2_{uv}\varphi\rangle + \int_0^t \langle \frac{1}{2}(vol_s^{\tilde{X}})^2\partial^2_{uu}C_s + \frac{1}{2}(vol_s^{\tilde{Y}})^2\partial^2_{vv}C_s + \tilde{\rho}_s vol_s^{\tilde{X}}vol_s^{\tilde{Y}}\partial^2_{uv}C_s, \partial^2_{uv}\varphi\rangle ds$$

□

If there is more regularity, then a strong equation obtains:

Proposition 6 *Suppose $C_t \in \mathcal{C}^{1,4}(\mathbf{R}_+ \times ((0,1)^2))$ and that the marginal densities, the volatilities and the correlation are twice continuously differentiable in the space variables. Assume that $\partial_t C_t(u,v)$ goes to zero as either u or v goes to zero. Then, for all $t>0$ and $(u,v)\in(0,1)^2$,*

$$\begin{aligned}\partial_t C_t(u,v) =& \frac{1}{2}(vol_t^{\tilde{X}})^2\partial^2_{uu}C_t + \frac{1}{2}(vol_t^{\tilde{Y}})^2\partial^2_{vv}C_t + \tilde{\rho}_t vol_t^{\tilde{X}}vol_t^{\tilde{Y}}\partial^2_{uv}C_t \\ &+ \varphi_t(u) + \psi_t(v) + \alpha_t\end{aligned} \tag{3.8}$$

where $vol_t^{\tilde{X}}(u) = (f_t^X.\sigma_t^X)(t,(F_t^X)^{-1}(u))$ and $vol_t^{\tilde{Y}}(v) = (f_t^Y.\sigma_t^Y)(t,(F_t^Y)^{-1}(v))$. Defining $g_t(u,v)$ as the function on the right hand side of the first line, $\varphi_t(u) = -\lim_{\epsilon\to 0} g_t(u,\epsilon)$, $\psi_t(v) = -\lim_{\epsilon\to 0} g_t(\epsilon,v)$ and $\alpha_t = \lim_{\epsilon\to 0} g_t(\epsilon,\epsilon)$.

Proof: The proof is straightforward. The hypotheses of regularity of the copula and the coefficients imply that equation (3.7) is equivalent to the strong equation

$$\partial^2_{uv}\partial_t C_t = \partial^2_{uv}\left(\frac{1}{2}vol_t^{\tilde{X}}(u)^2\partial^2_{uu}C_t(u,v) + vol_t^{\tilde{Y}}(v)^2\partial^2_{vv}C_t(u,v) + \widetilde{\rho}_t(u,v)vol_t^{\tilde{X}}(u)vol_t^{\tilde{Y}}(v)\partial^2_{uv}C_t\right)$$

for all $t>0$, $(u,v)\in(0,1)^2$. Let $\epsilon>0$. For a bivariate function f, denote $\Delta_{a,b}^{c,d}f := f(b,d) - f(a,d) - f(b,c) + f(a,b)$. Integrating the previous equation between $\epsilon<u$ and $\epsilon<v$:

$$\Delta_{\epsilon,u}^{\epsilon,v}\partial_t C_t = \Delta_{\epsilon,u}^{\epsilon,v} g_t(u,v)$$

By hypothesis, the l.h.s. goes to $\partial_t C_t(u,v)$ when $\epsilon \to 0$ $\square$.

Thereafter, it is always assumed that the function $\varphi_t(u) + \psi_t(v) + \alpha_t$ is identically zero, and therefore that for all $t > 0$, for all $(u,v) \in (0,1)^2$,

$$\begin{aligned}\partial_t C_t(u,v) = \frac{1}{2}\Big(\Big(\widetilde{f_t^X \sigma^X}(t,u)\Big)^2 \partial^2_{uu} C_t(u,v) + \Big(\widetilde{f_t^Y \sigma^Y}(t,v)\Big)^2 \partial^2_{vv} C_t(u,v)\Big) \\ + (\rho_t f_t^X(u)\sigma^X \widetilde{(t,u) f_t^Y}(v)\sigma^Y(t,v))\partial^2_{uv} C_t(u,v)\end{aligned} \tag{3.9}$$

In what follows, only this equation is referred to as the 'copula PDE' (rather than the weak equation (3.7)). Nevertheless, here are sufficient conditions for the extra terms in (3.8) to vanish :

1. For all $t > 0$, the marginal densities f_t^X, f_t^Y and the volatilities $\sigma^X(t,\cdot)$, $\sigma^Y(t,\cdot)$ are bounded over $\mathbf{R}$.

2. For all $t > 0$, for all $(u,v) \in (0,1)^2$,

$$\lim_{\epsilon \to 0} \partial^2_{uu} C_t(u,\epsilon) = \lim_{\epsilon \to 0} \partial^2_{vv} C_t(\epsilon,v) = \lim_{\epsilon \to 0} \partial^2_{uv} C_t(u,\epsilon) = \lim_{\epsilon \to 0} \partial^2_{uv} C_t(\epsilon,v) = 0$$

3. For all $t > 0$, for all $(u,v) \in (0,1)^2$,

$$\lim_{\epsilon \to 0} vol_t^{\tilde{X}}(\epsilon)^2 \partial^2_{uu} C_t(\epsilon,v) = \lim_{\epsilon \to 0} \partial^2_{vv} vol_t^{\tilde{Y}}(\epsilon)^2 C_t(u,\epsilon) = 0$$

The conditions 1 and 2 are satisfied in the rest of this chapter, when the marginals are Brownian motions. Our experiments show it is also the case of condition 3, although it is more difficult to prove rigorously that it holds for a given copula family $\{C_t\}$.

Finally, let us mention that proposition 6 still holds when the initial distribution μ_0 is singular:

Corollary 1 *Suppose that the coupled diffusion has a smooth copula $C_t \in \mathcal{C}^{1,2}(\boldsymbol{R}_+^* \times (0,1)^2)$, Assume the cdfs of the marginals are twice continuously differentiable in space for all $t > 0$ and continuously differentiable in the time variable on $\boldsymbol{R}_+^*$, and the same for the volatilities. Then equation (3.9) holds for all $t > 0$.*

This result is obtained by considering the time-shifted SDE (3.6), see appendix 3.6.1.

Uniqueness

The explicit expression between the coupling correlation and the copula family $\{C_t\}$ suggests considering the coupling problem from the opposite point of view; namely, if a copula family $\{C_t\}$ is fixed, that the function ρ_t is defined according to equation (3.9) is bounded by 1 and yields a solution to the coupling equation (3.6), then it is sensible to suppose that the resulting

diffusion process (X_t, Y_t) has copula C_t at time t. Such a result follows from the uniqueness of solutions to Kolmogorov forward equation. Let **(IC)** denote the integrability condition:

$$\exists\, \varepsilon > 0 \ s.t. \ (u,v) \mapsto \sup_{0<t\leq\varepsilon} \partial^2_{uv} C_t(u,v) \in L^1((0,1)^2)$$

Proposition 7 *Assume the marginal diffusions have densities f^X_t and f^Y_t in $\mathcal{C}^1(\mathbf{R}^*_+ \times \boldsymbol{R})$ for all $t > 0$, and that they and their derivatives goes to zero at $\pm\infty$. Let $\{C_t\} \in \mathcal{C}^{1,2}(\mathbf{R}_+ \times (0,1)^2)$ satisfying **(IC)**. Let $\rho(t,u,v)$ such that the copula PDE (3.9) holds, and $\rho(t,u,v)$ is bounded by one in absolute value. Assume that the drift and volatilities a^X, a^Y, σ^X, σ^Y, and the correlation function $\tilde{\rho}_t(x,y) = \rho(t, (F^X_t)^{-1}(u), (F^Y_t)^{-1}(v))$ are jointly continuous over $\boldsymbol{R}_+ \times \boldsymbol{R}^2$, measurable and bounded. Assume eventually that ρ is bounded away from ± 1, uniformly in time, i.e. $\inf_{t>0,\, x\, y} ||\rho(t,x,y)| - 1| \geq \varepsilon$ for some $\varepsilon > 0$. Then if the coupling SDE admits a solution, its law has a density at all times $t > 0$ and the copula at time t is indeed C_t.*

See appendix 3.6.1 for a detailed proof.

Note that $\tilde{\rho}_t(x,y) = \rho(t, F^X_t(x), F^Y_t(y))$ might not make sense at time $t = 0$ when the marginal distributions are singular at $t = 0$, unless $\rho(t,u,v)$ is a constant for $t < \varepsilon$. This is the case for instance when C_t is the Gaussian copula with constant parameter for $t < \varepsilon$ (and in this case the condition **(IC)** is also satisfied) and that the marginals diffusions are Brownian motions or geometric Brownian motions.

3.3 The case of coupled Brownian motions

While the previous section detailed the link between the copula of a coupled diffusion and the coupling correlation, it remains unclear whether a given copula family $\{C_t\}$ yields a function ρ_t that is indeed a correlation function, that is, at least, a function bounded by 1. This section is devoted to the coupling problem when the marginals are Brownian motions. A detailed example shows that it is possible to couple Brownian motions in a 'stationary' manner by a non Gaussian copula. Several examples of copulas that yield admissible coupling correlation functions as well as counterexamples are mentioned and an heuristic characterization of copulas that are attainable by coupled Brownian motions is discussed.

3.3.1 The coupling problem when marginals are Brownian motions

When the marginal processes are Brownian motions, the coupling SDE (3.6) becomes :

$$\begin{cases} dB^2_t &= \rho_t(\Phi(B^1_t/\sqrt{t}), \Phi(B^2_t/\sqrt{t}))dB^1_t + \sqrt{1 - \rho^2_t(\Phi(B^1_t/\sqrt{t}), \Phi(B^2_t/\sqrt{t}))}dZ_t \\ B^2_0 &= 0 \end{cases} \tag{3.10}$$

where (B_t^1, Z_t) is a standard bivariate Brownian motion, Φ is the cdf of the standard normal law and $(t,u,v) \mapsto \rho(t,u,v)$ is in $\mathcal{C}^0(\mathbf{R}_+ \times (0,1)^2)$.

Let Corr be the set of functions ρ such that

1. $\rho \in \mathcal{C}^0(\mathbf{R}_+ \times (0,1)^2)$.
2. $\sup_{(t,u,v)} |\rho(t,u,v)| \leq 1$.
3. ρ is constant for t small enough, i.e. there exists $\delta > 0$ such that $\rho(t,\cdot) = \rho \in [-1,1]$ for all $t \leq \delta$.
4. ρ is bounded away from ± 1: $\exists\, \varepsilon > 0$, s.t. for all (t,u,v), $||\rho(t,u,v)| - 1| > \varepsilon$.

The conditions 3 and 4 ensure that the correlation function $\rho\Big(t, \frac{\Phi(x)}{\sqrt{t}}, \frac{\Phi(y)}{\sqrt{t}}\Big)$ make the coupling SDE have a unique strong solution (see proposition 17 in appendix 3.6.2). In particular, the problem of defining the quantiles $(F_t^X)^{-1}$ and $(F_t^Y)^{-1}$ at $t = 0$ is avoided, thanks to condition 3, which imposes a constant Gaussian copula at small times. Likewise, the set of continuous functions ρ defined on $(0,1)^2$ such that $|\rho| \leq 1$ and ρ is bounded away from ± 1 is denoted $\mathrm{Corr}_{\mathrm{BM}}$ (for Brownian correlation function).

Let $\mathcal{C}^+ = \{\{C_t\}_{t\geq 0},\ C_t \in \mathcal{C}^{1,2}(\mathbf{R}^+ \times (0,1)^2),\ \partial^2_{uv} C_t(u,v) > 0,\ \forall (u,v) \in (0,1)^2\}$, the smooth copula families with everywhere positive densities. It is convenient to consider the mapping:

$$F : \{C_t\} \in \mathcal{C}^+ \mapsto \rho_{C_t} \in \mathcal{C}^0(\mathbf{R}_+ \times (0,1)^2)$$

$$\text{where } \rho_{C_t} : (u,v) \in (0,1)^2 \mapsto \frac{2\pi t \partial_t C_t(u,v) - \frac{1}{2}\left(e^{-\Phi^{-1}(u)^2}\partial^2_{uu} C_t(u,v) + e^{-\Phi^{-1}(v)^2}\partial^2_{vv} C_t(u,v)\right)}{e^{-\frac{\Phi^{-1}(u)^2 + \Phi^{-1}(v)^2}{2}}\partial^2_{uv} C_t(u,v)} \tag{3.11}$$

$F(\{C_t\})$ is the correlation function that appears in the copula PDE (3.9), when the copula family is $\{C_t\}$. Conversely, it is convenient to consider

$$G : \rho_t \in \mathrm{Corr} \mapsto \{C_t\},\ \text{the copula family of the coupled BMs with correlation function } \rho_t$$

Let $\mathrm{Cop} = G(\mathrm{Corr})$ be the set of copula families that are the copula family of coupled Brownian motions with a correlation function in Corr. For $\rho \in \mathbf{R}$, $|\rho| \leq 1$, $\mathrm{Corr}(\rho)$ denotes the set of $\rho_t \in \mathrm{Corr}$ such that $\rho_t = \rho$ for all t small enough: $\mathrm{Corr} = \cup_{\rho \in (-1,1)} \mathrm{Corr}(\rho)$. Finally, let $\mathrm{Cop}(\rho) = F^{-1}(\mathrm{Corr}(\rho))$, the set of copulas such that the correlation function is in $\mathrm{Corr}(\rho)$.

A question of crucial importance is to determine the set of copula that can be attained by coupled Brownian motions, that is to say the copulas C such that there exists a coupled Brownian motion (i.e. a correlation function) whose copula C_T at some time T satisfy $C_T = C$. More precisely,

Definition 4 *A copula C is* attainable *at time $T \geq 0$, iff there exists $\{C_t\} \in Cop$ such that $C_T = C$. The set of attainable copulas at some time $t \geq T$ is denoted A_T. A copula C is said* stationary *(for the Brownian motion) if it is attainable at some time $T \geq 0$ by coupled Brownian motions (B_t^1, B_t^2), and that $C_{(B_t^1, B_t^2)} \equiv C$, for all $t \geq T$. The set of stationary copulas from time T is denoted A_T^S.*

A trivial example of a stationary is the Gaussian copula, but it is not the only copula to have this property, as is proved below. Along with the notion of stationary copulas comes the notion of stationary (Brownian) correlation function:

Definition 5 *For $C \in \mathcal{C}^2((0,1)^2)$ a copula with positive density, let the* stationary Brownian correlation function of C *be the function*

$$\rho_C(u,v) = -\frac{1}{2}\frac{e^{\frac{\Phi^{-1}(v)^2 - \Phi^{-1}(u)^2}{2}}\partial^2_{uu}C(u,v) + e^{\frac{\Phi^{-1}(u)^2 - \Phi^{-1}(v)^2}{2}}\partial^2_{vv}C(u,v)}{\partial^2_{uv}C(u,v)}, \forall (u,v) \in (0,1)^2 \tag{3.12}$$

If a copula is stationary, then the correlation function ρ_t of the coupled Brownian motions which attain the dependence C is necessarily equal to ρ_C for t big enough. In particular, time does not appear in expression (3.12), and this expression is more convenient to work with than (3.11). Although it might seem trivial, the case of the Gaussian copula is worth noticing:

Proposition 8 *The stationary Brownian correlation function of the Gaussian copula C_ρ is constant over $(0,1)^2$ and equals ρ. Moreover, if $\rho_t \in \mathcal{C}^1(\boldsymbol{R}_+)$ and is bounded by 1, and that $C_t := C_{\frac{1}{t}\int_0^t \rho_s ds}$ if $t > 0$ and $C_0 = C_{\rho_0}$, then the correlation function $F(\{C_t\}_{t\geq 0})$ is ρ_t.*

See appendix 3.6.2 for a proof.

As expected, the stationary correlation of a Gaussian copula copula with constant parameter ρ is ρ.

Eventually, a set of particular interest is the intersection of all A_t, $t > 0$:

Definition 6 *Let $A_{0+} = \cap_{t>0} A_t$: it is the set of copulas that can be attained at arbitrary small times. Similarly, A_{0+}^S is defined as the set of stationary copulas from any time $t > 0$.*

The point is to prove that these sets are not reduced to the Gaussian copula family.

3.3.2 Results on the attainability of a copula C

The self-similarity property of Brownian motion considerably simplifies the analysis of the set of attainable copulas.

Lemma 5 *Let (B_t, W_t) a pair of coupled Brownian motions, with coupling correlation $\rho(t,x,y)$. Then for every $c > 0$, the bivariate process $(B_t^c, W_t^c) = (\frac{B_{ct}}{\sqrt{c}}, \frac{W_{ct}}{\sqrt{c}})$ is a coupled Brownian motion with correlation function $\rho^c(t,x,y) = \rho(ct, \sqrt{c}x, \sqrt{c}y)$. This correlation is in Corr if $\rho \in Corr$, and the copula of (B_t^c, W_t^c) is the copula of (B_{ct}, W_{ct}).*

Proof: By self-similarity of the Brownian motion, (B_t^c, W_t^c) is a process with Brownian marginals. Furthermore,

$$\langle B^c, W^c\rangle_t = \frac{1}{c}\langle B, W\rangle_{ct} = \frac{1}{c}\int_0^{ct} \rho(s, B_s, W_s)ds$$

and thus, $d\langle B^c, W^c\rangle_t = \rho^c(t, B_t^c, W_t^c)dt$ where $\rho^c(t, x, y) = \rho(ct, \sqrt{c}x, \sqrt{c}y)$. Eventually, the copula of (B_t^c, W_t^c) evaluated at $(u, v) \in [0, 1]^2$ is:

$$\mathbf{P}(\Phi(B_t^c/\sqrt{t}) \leq u, \Phi(W_t^c/\sqrt{t}) \leq v) = \mathbf{P}(\Phi(B_{ct}/\sqrt{ct}) \leq u, \Phi(W_{ct}/\sqrt{ct}) \leq v)$$

and the r.h.s. is, by definition, the copula of (B_{ct}, W_{ct}) evaluated at (u, v). The fact that $\rho^c \in$ Corr if $\rho \in$ Corr is obvious.□
A direct consequence is

Proposition 9 *$A_{0+} = \cup_{t>0} A_t$. In other words, if a copula is attainable at some time $T > 0$, then it is attainable at any time $t > 0$. Similarly $A_{0+}^S = \cup_{t>0} A_t^S$.*

Proof: Assume that $C \in A_T$ and let (B^1, B^2) a pair of coupled Brownian motions, such that $C_{B_T^1, B_T^2} = C$. Then, for $c > 0$, lemma 5 ensures that C is attainable a time $\frac{T}{c}$. □

In addition to self-similarity, the Brownian motion is stable by time-inversion, meaning that if B_t is a standard BM, then so is the process that starts at 0 at time 0 and is defined $tB_{\frac{1}{t}}$ if $t > 0$. Therefore,

Proposition 10 *Let (B, W) be a pair of coupled BMs. Then the copula of $(\tilde{B}_t, \tilde{W}_t) := \left(tB_{\frac{1}{t}}, tW_{\frac{1}{t}}\right)$ is the copula of $\left(B_{\frac{1}{t}}, W_{\frac{1}{t}}\right)$ for $t > 0$.*

Proof: For each $(u, v) \in [0, 1]^2$

$$\begin{aligned} C_{\tilde{B}_t, \tilde{W}_t}(u, v) &= \mathbf{P}(\Phi(\tilde{B}_t/\sqrt{t}) \leq u, \Phi(\tilde{W}_t/\sqrt{t}) \leq v) \\ &= \mathbf{P}\left(\Phi\left(\sqrt{t}B_{\frac{1}{t}}\right) \leq u, \Phi\left(\sqrt{t}W_{\frac{1}{t}}\right) \leq v\right) \\ &= C_{B_{\frac{1}{t}}, W_{\frac{1}{t}}}(u, v) \text{ by definition.} \end{aligned}$$

□
This process has Brownian marginals and proposition 10 shows that whenever a copula $C \in A_{0+}^S$, then there exists a pair of Brownian motions (A_t, B_t) such that the copula C_{A_t, B_t} is C for all t small enough. Note that the previous properties are specific to Brownian motion, and otherwise the sets A_{0+} might well be strictly included in A_t, $t > 0$.

A crucial point is to show that the set of stationary copulas is not reduced to the Gaussian family. In order to prove it, we first show that the sets Cop(ρ) have convexity properties.

Lemma 6 *For all $\rho \in (-1,1)$, $Cop(\rho)$ is stable by constant mixtures: if $\alpha \in [0,1]$, $\{C_t\}$ and $\{\tilde{C}_t\} \in Cop(\rho)$, then $\alpha C_t + (1-\alpha)\tilde{C}_t \in Cop(\rho)$.*

The proofs of the rest of this section are gathered in appendix 3.6.2. Note that the correlation function derived from the copula $\alpha_t C_\rho + (1-\alpha_t)C$ is not the convex sum $\alpha_t\rho + (1-\alpha_t)\rho_C(u,v)$, which complicates the analysis. Using time-dependent mixtures of Gaussian copulas and a given copula C, it is possible to prove that some copulas are stationary:

Proposition 11 *Let C be a copula such that $\rho_C \in Corr_{BM}$. Introduce*

$$\delta_C(\rho) = \inf_{\substack{(u,v)\in(0,1)^2 \\ s.t.\, C_t(u,v)\neq C(u,v)}} \left\{ \frac{e^{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}}{2\pi|C_\rho(u,v)-C(u,v)|}[(1-|\rho|)\partial^2_{uv}C_\rho(u,v) \wedge (1-|\rho_C(u,v)|)\partial^2_{uv}C] \right\}.$$

and assume that $\delta_C(\rho) > 2$ for some $\rho \in (-1,1)$. Then $C \in A^S_{0^+}$.

However, when a copula C has a stationary correlation function ρ_C in $\mathrm{Corr}_{\mathrm{BM}}$ but does not satisfy $\delta_C(\rho) \leq 2$, our intuition is that it can be attained, at least asymptotically. It means that defining the convex sum $\rho_t(u,v) = \alpha_t\rho + (1-\alpha_t)\rho_C(u,v)$ for a convenient function α_t, we expect the resulting coupled Brownian motions to have a copula C_t such that $C_t \to C$ as $t \to \infty$. This convergence has been observed empirically on simulations of coupled Brownian motions for various correlation functions ρ: after a few steps of an Euler scheme, the copula seems to stop evolving anymore and becomes stationary. Eventually, this intuition is supported by the fact that such a copula satisfies the equation $(\mathcal{L}_t^{scaled})^*\partial^2_{uv}C = 0$ for t big enough, and it is a stylized fact in the theory of Markov processes that this indicates that C is a possible stationary distribution.

3.3.3 A detailed example: the FGM copula

The FGM copula (Farlie-Gumbel-Morgenstern copula) is defined by $C_\theta(u,v) = uv + \theta uv(1-u)(1-v)$, for $|\theta| \leq 1$. This copula family contains all copulas with quadratic sections in both u and v, i.e. all copulas C such that both $C(u,\cdot)$ and $C(\cdot,v)$ are quadratic functions. This type of copula does not produce strong dependence: as $\partial_\theta C(u,v) \geq 0$, it is positively ordered family, and Nelsen [15] p. 78 provides scatterplots from the extremal members C_{-1} and C_1 which clearly exhibit a low degree of dependence. Notwithstanding its relevance to model strong dependence, this family has the advantage of yielding particularly simple formulas that allow for explicit computations. In particular, the stationary correlation function of the FGM copula C_θ is

$$\rho_\theta(u,v) = \theta.\frac{e^{(\Phi^{-1}(v)^2-\Phi^{-1}(u)^2)/2}v(1-v) + e^{(\Phi^{-1}(u)^2-\Phi^{-1}(v)^2)/2}u(1-u)}{1+\theta(1-2u)(1-2v)} \tag{3.13}$$

Proposition 12 *For all* $|\theta| \leq \frac{1}{2}$, $|\rho_\theta(u,v)| \leq 1$ *for all* u, v *in* $[0,1]$.

Proof: : c.f. appendix 3.6.2. A numerical analysis of ρ_θ suggests the sharp bound $|\rho_\theta(u,v)| \leq \frac{|\theta|}{2}$ holds for all $|\theta| \leq 1$ and we will use this bound in what follows to prove that the FGM copula is a stationary copula.

Proposition 13 *There exists a non empty range of parameters* $\theta \in [-\alpha, \alpha]$ *, such that the FGM copula* $C_\theta \in A^S_{0+}$.

Proof: Apply the proposition 11 with $\rho = 0$ and $\theta \neq 0$ (which corresponds to the independence copula). Then we have:

$$\inf_{(u,v)\in(0,1)^2} \frac{e^{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}}{2\pi(C_\theta(u,v)-uv)}.\left((1-|\rho_\theta|).\partial^2_{uv}C_\theta \wedge 1\right) \geq \frac{16(1-|\theta|)(1-|\theta|/2)}{2\pi|\theta|} \tag{3.14}$$

Indeed, $|\rho_\theta| \leq \frac{|\theta|}{2}$ and $\partial^2_{uv}C_\theta = 1+\theta(1-2u)(1-2v) \geq 1-|\theta|$, so $(1-\rho_\theta).\partial^2_{uv}C_\theta \wedge 1 \geq 1 \wedge (1-|\theta|)(1-|\theta|/2) = (1-|\theta|)(1-|\theta|/2)$. Eventually $C_\theta(u,v)-uv = |\theta|uv(1-u)(1-v) \leq |\theta|/16$ gives the inequality. A sufficient condition for the proposition 11 to apply is then $\frac{16(1-|\theta|)(1-|\theta|/2)}{2\pi|\theta|} > 2$, which is easily seen to be true for all $0 \leq |\theta| \leq \alpha$, $\alpha \approx 0.49$. □
The lower bound (3.14) is not a sharp one and actually, numerical evidence suggests that the whole FGM copula family is in A^S_{0+}.

This result has some importance, as it proves that there exists bivariate processes, whose marginals are Brownian motions, and which are coupled by a non Gaussian copula , from an arbitrary small time $t > 0$:

Corollary 2 A^S_{0+} *does not reduce to Gaussian copulas and contains members of the FGM family.*

3.3.4 Stationary copulas of some processes with Gaussian marginals

Back to the more general case of coupling Markovian diffusions, consider the problem of coupling Markovian diffusions that have the same marginals, and that these marginals have Gaussian 1 dimensional laws, i.e. two diffusions X_t and Y_t with same drift and volatility, such that for all $t > 0$, X_t and Y_t have the same Gaussian law $\mathcal{N}(m_t, \sigma_t^2)$. Suppose moreover the volatility of the marginals is a deterministic function of time $\Sigma_t > 0$. Then (assuming that the equation (3.9) holds and the density of the copula is everywhere positive) the coupling correlation reads

$$\rho(t,u,v) = 2\pi.e^{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}} \frac{\sigma_t^2}{\Sigma_t^2} \frac{\partial_t C_t}{\partial^2_{uv} C_t} + \rho_C(u,v)$$

because $f^X_t \circ (F^X_t)^{-1}(u)$ is proportional to $e^{-\Phi^{-1}(u)^2/2}$, that is, the scaled densities are proportional to the scaled $\mathcal{N}(0,1)$ density, just as in the case of the Brownian motion. In particular, if

the copula is stationary, then there exists a $T > 0$ such that for all $t \geq T$, $\rho(t, u, v) = \rho_C(u, v)$, the stationary Brownian correlation function of the copula C. This applies for instance to Orstein-Uhlenbeck processes and to Brownian bridges.

3.3.5 A zoology of smooth copulas and their stationary correlations

Here are listed copulas for which we have explicit formulas for the stationary correlation function. Empirically, the copulas divide into two families: the one with a stationary correlation bounded by 1, and the one with a stationary correlation that explodes near the boundary of the unit square.

In addition to the Gaussian copula, numerical evidence suggests the following copulas have bounded correlation function:

- The FGM and the iterated FGM copulas (namely the Kotz and Johnson's and Li's iterated FGM copulas, that are families of copulas with cubic horizontal and vertical section, see [15], p 82).
- The Plackett copula C_θ, when $\theta \leq 10$.
- Among archimedean copulas: the Frank copula, the Gumbel-Barnett (and possibly others, for instance 'copula 4.2.10', see appendix 3.7).

Some of the corresponding Brownian correlation functions are plotted in figures 3.2 and 3.3 for the FGM and the Plackett correlation function (all figures are gathered at the end of the appendices). Note also that for a single copula C with positive density, such that $\sup_{[0,1]^2} |\rho_C| \leq 1$ or $\sup_{[0,1]^2} |\rho_C| > 1$, then three other copulas h ve the same property:

Proposition 14 *The Brownian correlation functions of the copulas* $C_1(u, v) = u - C(u, 1-v)$, $C_2(u, v) = v - C(1-u, v)$ *and* $C_3(u, v) = u + v - 1 + C(1-u, 1-v)$ *are* $\rho_{C_1}(u, v) = -\rho_C(u, 1-v)$, $\rho_{C_2}(u, v) = -\rho_C(1-u, v)$, $\rho_{C_3}(u, v) = \rho_C(1-u, 1-v)$. *Moreover* $1 - \sup_{[0,1]^2} |\rho_{C_i}|$ *has the same sign as* $1 - \sup_{[0,1]^2} |\rho_C|$, $i = 1, 2, 3$.

Proof: It is elementary. For instance,

$$\rho_{C_1}(u, v) = \frac{1}{2} \frac{e^{\frac{\Phi^{-1}(v)^2 - \Phi^{-1}(u)^2}{2}} \partial^2_{uu} C(u, 1-v) + e^{\frac{\Phi^{-1}(u)^2 - \Phi^{-1}(v)^2}{2}} \partial^2_{vv} C(u, 1-v)}{\partial^2_{uv} C(u, 1-v)}$$

Moreover, because the normal law is symmetric, $\Phi^{-1}(v) = -\Phi^{-1}(1-v)$ and the result follows $\square$.

If $C = C_{U,V}$ where (U, V) is a pair of uniform variable, then $C_1 = C_{U,1-V}$, $C_2 = C_{1-U,V}$ and $C_3 = C_{1-U,1-V}$ is the survival copula of (U, V).

On the other hand, many copulas have a stationary correlation with an explosive behavior near the boundary of the unit square (two of them are depicted in figure 3.4). For such copulas there does not exist coupled Brownian motions such that the copulas C_t are constant and equal to C on a non trivial time interval. Numerical evidence shows that many usual archimedean copulas fall in this case; here is a detailed proof for the Clayton copula:

Proposition 15 *The stationary Brownian correlation function of the Clayton copula with parameter $\theta > 0$ is unbounded on $(0,1)^2$.*

Proof: the stationary correlation function is (see appendix 3.7.3), is

$$\rho_\theta(u,v) = \frac{1}{2}\Big(e^{\frac{\Phi^{-1}(v)^2-\Phi^{-1}(u)^2}{2}}\frac{v}{u}(1-v^\theta) + e^{\frac{\Phi^{-1}(u)^2-\Phi^{-1}(v)^2}{2}}\frac{u}{v}(1-u^\theta)\Big)$$

Fix $u \in (0,1)$; then $\lim_{v\to 0} corr_\theta(u,v) = +\infty$. Indeed, $\lim_{v\to 0} e^{\frac{-\Phi^{-1}(v)^2}{2}}/v = +\infty$, while $e^{\frac{\Phi^{-1}(v)^2}{2}}v(1-v^\theta)$ is bounded on $(0,\frac{1}{2}]$. This can be seen be writing that, for $x<0$,

$$e^{\frac{x^2}{2}}\Phi(x) = \int_{-\infty}^{0} e^{-u^2/2-xu}\frac{du}{\sqrt{2\pi}}$$

(see the proof of lemma 12). In particular, $0 \le e^{\frac{\Phi^{-1}(v)^2}{2}}v \le \frac{1}{2}$ when $v < \frac{1}{2}$, which, combined with $(1-v^\theta) \in [0,1]$ yields $e^{\frac{\Phi^{-1}(v)^2}{2}}v(1-v^\theta)$ is bounded by $\frac{1}{2}$. Finally, $\lim_{x\to-\infty} e^{\frac{x^2}{2}}\Phi(x) = 0$ by dominated convergence and thus $\lim_{v\to 0} e^{\frac{-\Phi^{-1}(v)^2}{2}}/v = +\infty$. $\square$.

Some elliptical copulas also have this property: the Student copula has an unbounded correlation function, and thus we claim the noticeable result that *it is not possible to couple Brownian motions from a fixed time on with a Student copula*, see figure 3.4.

The table 3.1 provides explicit formulas for the correlation function of the previously mentioned copulas (see also appendix 3.7). As the considered copulas are all symmetric, their stationary correlations read $\frac{a(u,v)+a(v,u)}{2}$ for some function $a(u,v)$, given in table 3.1. The copula is said 'Admissible' whenever its stationary correlation is bounded by 1.

There is an obvious difference between the copulas that are admissible and those that are not: the copulas with bounded stationary correlation we mentioned does not exhibit tail dependence, contrary to all the copulas which have unbounded correlation. Tail dependence measures the strength of the dependence of a copula in the lower-left quadrant and in the upper-right quadrant of $[0,1]^2$; for instance the coefficient of lower tail coefficient is defined as $\lim_{t\to 0^+}\frac{C(t,t)}{t}$. We refer to Nelsen [15] and to Jaworski's article [9] for precise statements on the tail dependence of the copulas mentioned above. Our experimental results leads us to

Copula	Stationary correlation function $(a(u,v))$	Admissible
Gaussian C^{ρ}	ρ	Yes
Student $C^{\rho,\nu}$	$\left(\rho + \frac{t_\nu^{-1}(u)t_\nu^{-1}(v)}{\nu}\right) e^{\frac{\Phi^{-1}(v)^2 - \Phi^{-1}(u)^2}{2}} \left(1 + \frac{t_\nu^{-1}(u)^2}{\nu}\right)^{\frac{\nu-1}{2}} \left(1 + \frac{t_\nu^{-1}(v)^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	No
Clayton C_θ	$e^{\frac{\Phi^{-1}(v)^2 - \Phi^{-1}(u)^2}{2}} \frac{v}{u}(1 - v^\theta)$	No
Gumbel C_θ	$-e^{\frac{\Phi^{-1}(v)^2 - \Phi^{-1}(u)^2}{2}} \frac{v}{u} \frac{(-\log(u))^{\theta-1}}{(-\log(v))^{\theta-1}} \frac{A^{1/\theta} - A(-\log(u))^{-\theta}(\theta - 1 - \log(u)) + \theta - 1}{A^{1/\theta} + \theta - 1}$ where $A = (-\log(u))^\theta + (-\log(v))^\theta$	No
Frank C_θ	$\frac{e^{-\theta}}{e^{-\theta}-1}(1 - e^{\theta v})(1 - e^{\theta(v-1)})$	Yes
Gumbel-Barnett C_θ	$\frac{\theta(1 - \theta \log(v)\frac{v}{u})}{1 - \theta - \theta\log(uv) + \theta^2 \log(u)\log(v)}$	Yes
FGM C_θ	$\frac{2\theta v(1-v)}{1+\theta(1-2u)(1-2v)}$	Yes
Plackett C_θ	$\frac{2(\theta-1)v(1-v)}{1+(\theta-1)(u+v-2uv)}$	Yes $(\theta < 10)$

Table 3.1: TO BE PUT

infer that stationary copulas are necessarily without tail dependence (although it may not be a sufficient condition for being a stationary copula).

3.3.6 A heuristic characterization of attainable copulas

The fact that the stationary correlation function ρ_C is not bounded by 1, indicates only that C is not a stationary copula, but does not prove that it is not attainable. Nevertheless, we outline here a heuristic characterization of the distributions that are attainable by coupled Markovian diffusions found by A. Galichon ([8]). The idea is to write a variational problem consisting in minimizing an objective function which depends on the correlation, and then formally writing the dual problem; this latter problem has the property to have finite value iff the copula is attainable at the chosen horizon.

Proposition 16 (*Heuristic*) *Let p and q be two probability distributions over $\mathbf{R}$, and f a function defined on $[-1,1]$. Consider the problem*

$$\inf_{\substack{p_t,\rho_t \ s.t.\\ p_0=p,p_T=q\\ \partial_t p_t-\frac{1}{2}\triangle p_t-\partial^2_{xy}(\rho p_t)=0}} \int\int_0^T L^f(\rho_t(x))p_t(x)dtdx \tag{3.15}$$

where $L^f(x)=\mathbb{1}_{|x|\leq 1}f(x)+(+\infty)\mathbb{1}_{|x|>1}$. Then the primal problem (3.15) admits the following dual

$$S:=\sup_{\substack{\varphi_t \ s.t.\\ \partial_t\varphi_t+\frac{1}{2}\triangle\varphi_t+f^*(\partial^2_{xy}\varphi_t)=0}} \int\varphi(T,x)p_T(x)dx-\int\varphi(0,x)p_0(x)dx \tag{3.16}$$

where $f^*(y)=\sup_{|x|\leq 1}(x\cdot y-f(x))$ is the Legendre transform of f over $[-1,1]$.
The PDE which appears in the constraints of the primal problem is the Kolmogorov forward equation of the coupled Brownian motion with correlation function ρ, the distribution q is attainable at time T if and only if $S<\infty$.

Thus, theoretically, it should be possible to determine whether a given copula is attainable, by choosing a convenient function f and solving the problem (3.16) for every possible initial function φ_0. Eventually, remark that, as explained in section 3.3.2, the time parameter $T>0$ is not a decisive quantity in this problem, an S being finite is independent from its value.

3.4 A Financial example

In this section, a strategy of portfolio insurance is considered in order to assess the impact of coupling diffusions in practice. Portfolio insurance (and more precisely Constant Proportion Portfolio Insurance) are a class of dynamic strategies that aim at guaranteeing a protection at maturity (such as a nominal amount in the case of the classical CPPI) while benefiting of

the possible rise of a risky asset. These strategies fit well in our framework: they are dynamic and it is natural to use a continuous model of dependence to model the underlyings rather than imposing copulas at discrete times during the life of the strategy.

Definition and description of the strategy We opted for a particular type of strategy, called Long-Short CPPI (described in Amenc et al.[1] and Roncalli's book [19]), for the impact of dependence in such a strategy proved to be more obvious than for classical CPPI.
Let $T > 0$ be a time horizon. Let N be a nominal amount of cash. We assume the investment universe is composed of two risky assets, S_t^1 and S_t^2. The CPPI long-short strategy aims at guaranteeing a percentage α of the performance of the second asset, called the 'core' while benefiting from a possible rise of the first asset, the 'satellite'. Let $F_t = \frac{\alpha N}{S_0^2} S_t^2$ be the value at time t of the guarantee that must achieve the strategy. The dynamic CPPI strategy use leverage to invest in the satellite. It does it in in such a way that the value of the strategy always remains above F_t, possibly shorting the satellite and being long the core if it happens that $NAV_t < F_t$. The *multiplier* m is a real number that determines the strength of leverage; the higher m, the stronger the leverage. The *cushion* is equal to $C_t = NAV_t - F_t$ and the *cushion %* is $\frac{NAV_t - F_t}{NAV_t}$. The *investment level* is the proportion invested in the satellite: $IL_t = m.C_t/NAV_t$ meaning that the cushion is leveraged to invest in the satellite. It follows that, when the strategy is continuously rebalanced, the NAV has the dynamics:

$$\frac{dNAV_t}{NAV_t} = IL_t \frac{dS_t^1}{S_t^1} + (1 - IL_t)\frac{dS_t^2}{S_t^2}$$

Diffusion Model The chosen model is simple: the assets follow a coupled Black-Scholes model,

$$\begin{cases} \frac{dS_t^i}{S_t} & = \mu_t^i dt + \sigma_t^i dW_t^i, \ i = 1, 2 \\ d\langle W^1, W^2 \rangle_t & = \rho_t(W_t^1, W_t^2) dt \end{cases} \tag{3.17}$$

Thus the dynamics of the assets are log-normal, but the bivariate process (S_t^1, S_t^2) is not Gaussian in general. Moreover, if the assets are assumed to have the same dynamics, and the drifts are positive, then the copula of (S_t^1, S_t^2) is the same as the copula of (W_t^1, W_t^2), which further simplifies the analysis.

Results The impact of the dependence structure is seen on the *gap probability* at maturity. For such a strategy, the gap risk at maturity is the possibility that the value of the strategy at T is below F_T. It is the risk that the CPPI does not reach the level of protection and thus the risk that the CPPI seller suffers a loss. Of course, a gap can not occur if the CPPI is continuously rebalanced and the assets follow a continuous diffusion : the results obtained in this section are obtained when the strategy is rebalanced every 3 days, and the maturity is

one year. The two asset are assumed to be martingales ($\mu^i = 0$, $i = 1, 2$) and their volatilities are the same. In figure 3.4 are given an estimation of the gap probability $\mathbf{P}(NAV_T < F_T)$, for several copulas and several stock volatilities. In order to give a meaningful comparison of the copulas, the Sperman's rho is used as a measure of the strength of dependence. The Spearman's rho of a copula C is defined as the linear correlation between (U, V) of a pair of uniform random variable with copula C; it is a measure of concordance (whose definition is recalled in [15], p. 169), and is suitable to compare the strength of dependence across different copula families.

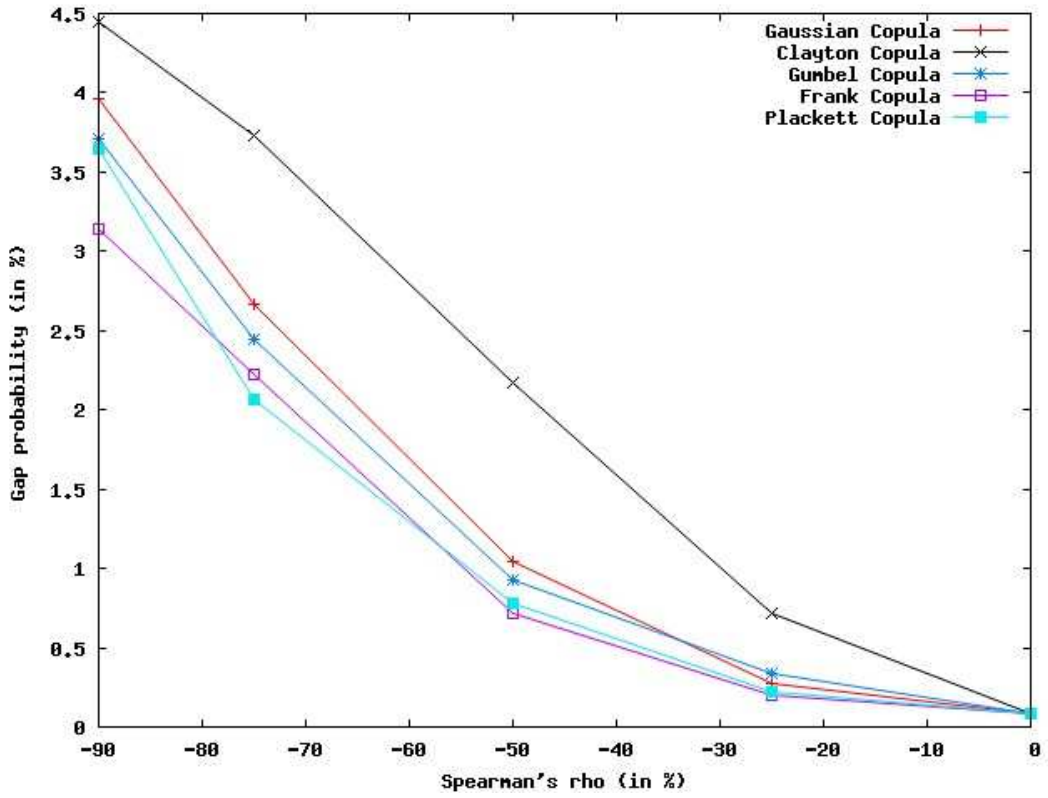

Figure 3.1: Gap probabilities obtained for different copulas.

Despite the fact that some of the considered copulas have unbounded stationary correlation ρ_C, we use nevertheless these copulas by forcing ρ_C to be bounded using the simplest possible 'trick', by considering $(u, v) \mapsto \mathbb{1}_{|\rho_C(u,v)| \leq 1} \rho_C(u, v) + sgn(\rho_C(u, v)) \mathbb{1}_{|\rho_C(u,v)|>1}$ (where $sgn(x)$ is the sign function, and equals 1 if $x \geq 0$, -1 otherwise). Of course, the copula of the resulting coupled Brownian motions is different from C [1]. The copula we consider all model

[1] However the results obtained in this manner are still labeled by the name original copula C in figure 3.4

'negative' dependence, as can be seen from their nonpositive Spearman's rho, and in particular, for copulas which exhibits positive dependence, such as the Gumbel copula, we use the 'C1' transform defined in proposition 14. This type of dependence is the 'adverse case' for Long-Short CPPI as it significantly increases the gap risk.

Figure 3.4 shows the gap probability as a function of the Spearman's rho for various copulas, when the volatility of both assets is 30%, that is in a market with high volatility. The impact of copulas is real, and in particular the gap probability obtained with the Clayton copula clearly bounds above the one obtained with the other copulas. This can be explained by the strong dependence generated by ('C1' transform of) Clayton copulas when the first asset drops while the second asset rises. It is noticeable that the Clayton copula is the only one to be 'uniformly' more conservative, i.e. yields higher gap probabilities, than the Gaussian copula.

3.5 Conclusion

This chapter addressed the problem of constructing bivariate continuous stochastic processes whose dependence at all times t is a given copula C_t, while the marginal processes are fixed Markovian diffusions. In particular, it tackles directly the problem of constructing Brownian motions whose cross-sectional dependence is controlled. It shows that while some of the most classic copulas can be used to model a stationary dependence between Brownian motions, it is nevertheless not the case for many of them, and we infer empirically that copulas which exhibit tail dependence might not be able to couple Brownian motions. These coupling models could be useful in stress testing and risk management of strategies, and we have provided first results as to the potential impact of dependence modeling with copulas on a long-short CPPI strategy.

This chapter develops the idea of coupling processes with copulas in a bivariate context only. The multivariate case (that is when more than two processes are involved) is more complex, and does not yield formulas as handy as in the bivariate case. While bivariate models allows for building non trivial multivariate models where the pairwise dependence is controlled (for instance star like dependence where the dependence between the processes (X_1, X_j) is imposed for all j, or serial dependence, for which the dependence between (X_{j-1}, X_j) is imposed), more work would be required to obtain general multivariate coupling models that would be exploitable in practice.

Finally, while we provided a case-by-case analysis of some copula families, a direction for further research is to obtain a general characterization of the set of copulas attainable by coupled Brownian motions, possibly by developing the ideas exposed in paragraph 3.3.6, that would be of practical interest to determine whether a given copula is an admissible model of dependence.

Bibliography

[1] N. Amenc, P. Malaise, and L. Martellini. Revisiting core-satellite investing. *Journal of Portfolio Management*, 31(1):64–75, 2004.

[2] A. Bentata and R. Cont. Mimicking the marginal distribution of a semimartingale. Preprint, 2011.

[3] U. Cherubini, F. Gobbi, S. Mulinacci, and S. Romagnoli. A copula-based model for spatial and temporal dependence of equity markets. In P. Jarowski, F. Durante, W. Härdle, and T. Rychlik, editors, *Copulas Theory and Its Applications, Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, volume 198 of *Lecture Notes in Statistics*. Springer Heidelberg Dordrecht London New-York, 2010.

[4] U. Cherubini and E. Luciano. Bivariate option pricing with copulas. *Appl. Math. Finance*, 9:69–82, 2002.

[5] W. Darsow, B. Nguyen, and E. Olsen. Copulas and Markov processes. *Ill. J. Math.*, 36(4):600–642, 1992.

[6] B. Dupire. Pricing with a smile. *Risk*, 1994.

[7] A. Galichon. Modelling correlation between two Markov diffusion processes: A copula approach, with application to stochastic correlation modelling. Technical Report, 2006.

[8] A. Galichon. A variational characterization of attainable copulas by coupled diffusion processes. Personal communication, 2012.

[9] P. Jaworski. Tail behavior of copulas. In P. Jarowski, F. Durante, W. Härdle, and T. Rychlik, editors, *Copulas Theory and Its Applications, Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*, volume 198 of *Lecture Notes in Statistics*. Springer Heidelberg Dordrecht London New-York, 2010.

[10] H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, 1997.

[11] N. V. Krylov and B. L. Rozovskii. Stochastic evolution equations. *Itogi Nauki i Tekhniki. Ser. Sovrem. Probl. Mat.*, 14:71–146, 1979.

[12] A. Langnau. A dynamic model for correlation. *Risk*, April 2010.

[13] D. X. Li. On default correlation: A copula function approach. *Journal of Fixed Income*, 9(4):43–54, 2000.

[14] T. Mikosch. Copulas: Tales and facts. *Extreme*, 9:3–20, 2006.

[15] R. B. Nelsen. *An introduction to copulas*. Springer Series in Statistics. Springer Science+Business Media, Inc., second edition edition, 2006.

[16] A. Patton. Modeling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556, May 2006.

[17] R. Plackett. A reduction formula for normal multivariate integrals. *Biometrika*, 41(3/4), December 1954.

[18] A. Reghai. Breaking correlation breaks. *Risk*, October 2010.

[19] T. Roncalli. *La Gestion d'Actifs Quantitative*. Finance. Economica, 2010.

[20] D. W. Stroock. *An Introduction to Partial Differential Equations for Probabilists*, volume 112 of *Cambridge Studies In Mathematics*. Cambridge University Press, 2008.

[21] D. W. Stroock and S. Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 1979.

[22] R. van den Goorbergh, C. Genest, and B. Werker. Bivariate option pricing using dynamic copula models. *Insur. Math. Econ.*, 37:101–114, 2005.

3.6 Appendix

3.6.1 Proofs for section 3.2

Proof of corollary 1 For $\varepsilon > 0$, consider the diffusion (3.6) shifted by ε. The shifted diffusion $(X_t^\varepsilon, Y_t^\varepsilon) = (X_{t+\varepsilon}, Y_{t+\varepsilon})$ is still a diffusion whose equation is:

$$\left\{\begin{array}{rl} dX_t^\varepsilon &= a^X(t+\varepsilon, X_t^\varepsilon)dt + \sigma^X(t+\varepsilon, X_t^\varepsilon)dW_t^{X,\varepsilon} \\ dY_t^\varepsilon &= a^Y(t+\varepsilon, Y_t^\varepsilon)dt + \sigma^Y(t+\varepsilon, Y_t^\varepsilon)dW_t^{Y,\varepsilon} \\ d\langle W^{X,\varepsilon}, W^{Y,\varepsilon}\rangle_t &= \rho_t(X_t^\varepsilon, Y_t^\varepsilon)dt \\ (X_0^\varepsilon, Y_0^\varepsilon) &\sim (X_\varepsilon, Y_\varepsilon) \end{array}\right. \tag{3.18}$$

By assumption the cdfs of the marginals, $F^X_{t+\varepsilon}$ and $F^Y_{t+\varepsilon}$ satisfy the hypotheses of proposition 5 (they are regular up to time 0). Thus the copula PDE is valid for the shifted diffusion and reads, for all $t > 0$, for all $\varphi \in \mathcal{C}_c^2((0,1)^2)$,

$$\begin{aligned}\langle(\partial_t C)_{t+\varepsilon}, \partial^2_{uv}\varphi\rangle = \Big\langle \frac{1}{2}\Big(&\Big(\widetilde{f^X_{t+\varepsilon}\sigma^X}(t+\varepsilon,\cdot)\Big)^2\partial^2_{uu}C_{t+\varepsilon} + \Big(\widetilde{f^Y_{t+\varepsilon}\sigma^Y}(t+\varepsilon,\cdot)\Big)^2\partial^2_{vv}C_{t+\varepsilon}\Big) \\ &+ (\rho_{t+\varepsilon}f^X_{t+\varepsilon}(\cdot)\sigma^X\widetilde{(t+\varepsilon,\cdot)f^Y_{t+\varepsilon}}(\cdot)\sigma^Y(t+\varepsilon,\cdot))\partial^2_{uv}C_{t+\varepsilon}, \partial^2_{uv}\varphi\Big\rangle\end{aligned} \tag{3.19}$$

and this is true for all $\varepsilon > 0$. □

Proof of proposition 7 The two next lemmas are needed to handle the initial singular distribution:

Lemma 7 *Let F_t be the cdf of X_t, where X_t is a continuous process such that $\lim_{t\to 0} X_t = x_0$ a.s. Then for all $q \in (0,1)$, $\lim_{t\to 0} F_t^{-1}(q) = x_0$.*

Proof: Suppose it is not the case. Then, there exists some $\varepsilon > 0$, and a sequence t_n such that $\lim_{n\to\infty} t_n = 0$ and $|F_{t_n}^{-1}(q) - x_0| > \varepsilon$ for all n. Remark that for all $x \in \mathbf{R}$, $\lim_{t\to 0} F_t(x) = \mathbb{1}_{x\geq x_0}$ by dominated convergence. Thus $F_{t_n}(x_0) \underset{n\to\infty}{\longrightarrow} 1$ and thus, for all n big enough, $F_{t_n}(x_0) > q$. By definition of the quantile function [2], $F_{t_n}^{-1}(q) \leq x_0$. Combined with $|F_{t_n}^{-1}(q) - x_0| > \varepsilon$, it yields $F_{t_n}^{-1}(q) \leq x_0 - \varepsilon$. As F_{t_n} is nondecreasing, and using $F_{t_n} \circ (F_{t_n}(x))^{-1} \geq Id$, one gets $q \leq F_{t_n}(x_0 - \varepsilon)$. The r.h.s converges to 0 as $n \to \infty$ and there is a contradiction. □

Lemma 8 *Let $\{C_t\}$ be a copula family in $\mathcal{C}^{1,2}(\boldsymbol{R}_+ \times (0,1)^2)$ and F_t^X and F_t^Y be the cdfs of continuous processes X_t and Y_t such that $\lim_{t\to 0} X_t = x_0$ a.s. and $\lim_{t\to 0} Y_t = y_0$ a.s. Suppose F_t^X and F_t^Y have positive derivatives for all $t > 0$. Assume the technical condition*

[2] For all $q \in [0,1]$, $F^{-1}(q) = \inf\{x | F(x) > q\}$

(IC): *$\exists \varepsilon > 0$ s.t. $(u,v) \mapsto \sup_{0<t\leq\varepsilon} \partial^2_{uv} C_t(u,v)$ is integrable over $(0,1)^2$. Let $p_t(x,y) = \partial^2_{xy}(C_t(F_t(x), F_t^Y(y)))$; then $\lim_{t\to 0} p_t = \delta_{(x_0,y_0)}$ in distribution, i.e. $\lim_{t\to 0} \mathbf{E}_{p_t}(\varphi(X_t,Y_t)) = \varphi(x_0,y_0)$ for all φ continuous and bounded.*

Proof:

$$\begin{aligned}\mathbf{E}_{p_t}(\varphi(X_t,Y_t)) &= \mathbf{E}(\varphi((F_t^X)^{-1}(U_t),(F_t^Y)^{-1}(V_t))), \text{ where } (U_t,V_t) \sim \partial^2_{uv} C_t \\ &= \int_{(0,1^2)} \varphi((F_t^X)^{-1}(u),(F_t^Y)^{-1}(v))\partial^2_{uv} C_t(u,v)dudv\end{aligned}$$

According to lemma 7, $\lim_{t\to 0} \varphi((F_t^X)^{-1}(u),(F_t^Y)^{-1}(v)) = \varphi(x_0,y_0)$. Eventually, the technical condition ensures that the dominated convergence theorem can be applied, yielding the result. □

Lemma 9 *Assume $f_t^X \in \mathcal{C}^{1,2}(\boldsymbol{R}_+^* \times \boldsymbol{R})$, $a_t^X \in \mathcal{C}^{0,1}(\boldsymbol{R}_+ \times \boldsymbol{R})$ is bounded, $\sigma_t^X \in \mathcal{C}^{0,2}(\boldsymbol{R}_+ \times \boldsymbol{R})$ and σ_t, σ_t' are bounded and that f_t and f_t' goes to 0 at $-\infty$. Then the cdf $F^X(x) = \int_{-\infty}^x f_t^X(z)dz$ satisfies, for all $t > 0$, for all $x \in \boldsymbol{R}$,*

$$\partial_t F_t = -a_t^X(x) f_t^X(x) + \frac{1}{2}\partial_x\{\sigma_t^2 f_t^X\}(x)$$

Proof: The proof consists in writing down the Kolmogorov forward equation for the density f_t^X and then summing from $-\infty$ to x. The boundary terms vanish by hypothesis.□

Before proving the proposition 7, let us recall a theorem that ensures the uniqueness in Kolmogorov forward equation:

Theorem 5 (Bentata, Cont [2]) *Suppose that:*

1. *The drifts, volatilities and correlation are measurable and bounded.*
2. *The drifts, volatilities and correlation are continuous in x, uniformly over the compacts in t*[3].
3. *The covariance matrix is coercive: $\forall R > 0$, $\forall t$, $\inf_{|z|\leq R} \inf_x x' a(t,z) x > 0$.*

Then for all x_0, there exists a unique family $p_t(x_0, dy)$ of probability measures with $p_0(x_0,\cdot) = \delta_{x_0}$ and for all $g \in \mathcal{C}_c^\infty(\boldsymbol{R}^2)$,

$$\int g(y)\partial_t p_t(x_0,dy) = \int \mathcal{L}_t g(y) p_t(x_0,dy)$$

[3] $f(t,x)$ continuous in x, uniformly over the compacts in t means that $\forall x$, $\forall T$ $\inf_{\delta>0} \sup_{t\in[0,T]} \inf_{|x'-x|\leq\delta} |f(t,x) - f(t,x')| = 0$.

Proof of prop 7: Let P_t be the law of probability with copula C_t and marginals densities f_t^X and f_t^Y. We aim at proving that, necessarily, P_t is the law at time t of the diffusion under consideration. This is done by proving that P_t satisfies the same Kolmogorov forward equation as the law of the diffusion, and then use theorem 5 to prove the uniqueness of the solutions of this equation.
For each $t > 0$, let $Fr(t, x, y)$ be the cdf of P_t, i.e. $Fr(t, x, y) = C_t(F_t^X(x), F_t^Y(y))$. $F_t(x, y)$ denotes the vector $(F_t^X(x), F_t^Y(y))$. Then, for all $t > 0$, x, $y \in \mathbf{R}$,

$$\partial_t Fr(t,x,y) = \partial_t C_t \circ F_t(x,y) + \partial_u C_t \circ F_t(x,y)\partial_t F_t^X(x) + \partial_v C_t \circ F_t(x,y)\partial_t F_t^Y(y) \qquad (3.20)$$

Now,

$$\begin{aligned}\partial_t C_t \circ F_t(x,y) = &\frac{1}{2}(\sigma_t^X(x) f_t^X(x))^2 \partial_{uu} C_t \circ F_t(x,y) + \frac{1}{2}(\sigma_t^Y(y) f_t^Y(y))^2 \partial_{vv} C_t \circ F_t(x,y) \\ &+ \rho_t(x,y)\sigma_t^X(x)\sigma_t^Y(y) f_t^X(x) f_t^Y(y) \partial_{uv}^2 C_t \circ F_t(x,y)\end{aligned}$$

according to the PDE. According to lemma 9, the marginal cdfs satisfy, for all $t > 0$, for all $x \in \mathbf{R}$,

$$\partial_t F_t^X = -a_t^X(x) f_t^X(x) + \frac{1}{2}\partial_x(\sigma_t^2(t,x) f_t(x))$$

and thus

$$\partial_u C_t \circ F_t(x,y)\partial_t F_t^X(x) = \{-a_t^X(x) f_t^X(x) + \frac{1}{2}\partial_x(\sigma_t^2(t,x) f_t^X(x))\}\partial_u C_t \circ F_t(x,y)$$

Remark that $\partial_x Fr(t,x,y) = f_t^X(x)\partial_u C_t \circ F_t(x,y)$, $\partial_{xx}^2 Fr(t,x,y) = f_t^X(x)^2 \partial_{uu}^2 C_t \circ F_t(x,y) + (f_t^X)'\partial_u C_t \circ F_t(x,y)$ and $\partial_{xx}^2 Fr(t,x,y) = f_t^X(x) f_t^Y(y)\partial_{uv}^2 C_t \circ F_t(x,y)$. Therefore (3.20) reads

$$\begin{aligned}\partial_t Fr(t,x,y) = &\rho_t(x,y)\sigma_t^X(x)\sigma_t^Y(y)\partial_{xy}^2 Fr(t,x,y) \\ &+ \frac{1}{2}(\sigma_t^X(x))^2 \partial_{xx}^2 Fr(t,x,y) - \frac{1}{2}(\sigma_t^X(x))^2 (f_t^X(x))'\partial_u C_t \circ F_t(x,y) \\ &+ \frac{1}{2}(\sigma_t^Y(y))^2 \partial_{yy}^2 Fr(t,x,y) - \frac{1}{2}(\sigma_t^X(y))^2 (f_t^Y(y))'\partial_v C_t \circ F_t(x,y) \\ &+ \{-a_t^X(x) f_t^X(x) + \frac{1}{2}\partial_x(\sigma_t^2(t,x) f_t^X(x))\}\partial_u C_t \circ F_t(x,y) \\ &+ \{-a_t^Y(y) f_t^Y(y) + \frac{1}{2}\partial_y(\sigma_t^2(t,y) f_t^Y(y))\}\partial_v C_t \circ F_t(x,y)\end{aligned}$$

However, expanding $\partial_x(\sigma_t^2(t,x) f_t^X(x))$, the terms in front of $\partial_u C_t \circ F_t(x,y)$ reduce to

$$\begin{aligned}&\left(-a_t^X(x) + \frac{1}{2}\partial_x\left(\sigma_t^X(x)^2\right)\right) f_t^X(x)\partial_u C_t \circ F_t(x,y) \\ &= \left(-a_t^X(x) + \frac{1}{2}\partial_x\left(\sigma_t^X(x)^2\right)\right)\partial_x Fr(t,x,y)\end{aligned}$$

Finally,

$$\begin{aligned}\partial_t Fr(t,x,y) &= \rho_t(x,y)\sigma_t^X(x)\sigma_t^Y(y)\partial^2_{xy}Fr(t,x,y)\\ &+ \frac{1}{2}(\sigma_t^X(x))^2\partial^2_{xx}Fr(t,x,y) + \frac{1}{2}(\sigma_t^Y(y))^2\partial^2_{yy}Fr(t,x,y)\\ &+ \left(-a_t^X(x) + \frac{1}{2}\partial_x\Big(\sigma_t^X(x)^2\Big)\right)\partial_x Fr(t,x,y)\\ &+ \left(-a_t^Y(y) + \frac{1}{2}\partial_y\Big(\sigma_t^Y(y)^2\Big)\right)\partial_y Fr(t,x,y)\end{aligned}$$

If $\varphi \in \mathcal{C}_c^2(\mathbf{R}^2)$, the expression $\langle \partial_t Fr(t,x,y), \partial^2_{xy}\varphi\rangle$ yields, thanks to the previous equation and integration by parts in the sense of distributions:

$$\begin{aligned}\partial_t\langle f_t(x,y),\varphi\rangle &= \langle -\partial_x(a_t^X(x)f_t(x,y)) - \partial_y(a_t^Y(y)f_t(x,y)),\varphi\rangle\\ &+ \langle\frac{1}{2}\partial^2_{xx}(\sigma_t^X(x)^2 f_t(x,y)) + \frac{1}{2}\partial^2_{yy}(\sigma_t^Y(y)^2 f_t(x,y)),\varphi\rangle\\ &+ \langle\partial^2_{xy}(\rho_t(x,y)\sigma_t^X(x)\sigma_t^Y(y)\partial^2_{xy}f_t(x,y)),\varphi\rangle\end{aligned}$$

This is precisely the Kolmogorov forward equation for the coupled diffusion. To summarize: if C_t satisfy the copula PDE and $f_t(x,y)$ is a bivariate probability density defined by $\partial_{xy}\{C_t(F_t^X, F_t^Y)\}$ then f_t satisfies the Kolmogorov forward equation of the coupled diffusion equation.
In order to invoke the theorem 5, the degeneracy of f_t at time 0 must be handled with care. This is done by stating that for all $\varphi \in \mathcal{C}_c^2(\mathbf{R}^2)$, for all $t > \varepsilon$,

$$\langle f_t,\varphi\rangle = \langle f_\varepsilon,\varphi\rangle + \int_\varepsilon^t \langle f_s, \mathcal{L}_s(\varphi)\rangle ds \tag{3.21}$$

where $\mathcal{L}_s(\varphi) = \begin{pmatrix} a_t^X \\ a_t^Y \end{pmatrix}\cdot\nabla\varphi + Tr(\sigma\sigma^*(t,x,y)Hess(\varphi)))$ and $\sigma\sigma^*(t,x,y) = \begin{pmatrix} (\sigma_t^X)^2 & \rho_t\sigma_t^X\sigma_t^Y \\ \rho_t\sigma_t^X\sigma_t^Y & (\sigma_t^Y)^2 \end{pmatrix}$.
Eventually, according to lemma 8, $\langle f_t,\varphi\rangle \to \langle \delta_{(x_0,y_0)},\varphi\rangle$ as $t \to 0$ and the integrand in the r.h.s of (3.21) is an integrable function on $[0,t]$. Indeed, for all $s > 0$, $|\mathcal{L}_s(\varphi)(x,y)|$ is bounded on $[0,t]\times\mathbf{R}^2$ by a constant that depends only on t and the bounds on φ and its derivatives of order less than 2. Defining $f_0 = \delta_{(x_0,y_0)}$, $\langle f_s, \mathcal{L}_s(\varphi)\rangle$ is integrable over $[0,t]$ and letting $\varepsilon \to 0$,

$$\langle f_t,\varphi\rangle = \langle f_0,\varphi\rangle + \int_0^t \langle f_s, \mathcal{L}_s(\varphi)\rangle ds$$

This equation has a unique solution according to theorem 5. Thus f_t is indeed the law of the diffusion and, in particular, the copula of the diffusion is C_t, for all $t > 0$. □

3.6.2 Proofs for section 3.3

Proof of proposition 8: Consider the second part of the proposition and let $\tilde{\rho}_t = \frac{1}{t}\int_0^t \rho_s ds$. Notice that $t\partial C_t = (\rho_t - \tilde{\rho}_t)\partial_\rho C_{\tilde{\rho}_t}$. Moreover, according to the formulas given in appendix 3.7.1,

$$\begin{aligned}\partial_\rho C_\rho &= \tfrac{1}{2\pi} e^{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}} \partial^2_{uv} C_\rho \\ \partial^2_{uu} C_\rho &= -\rho . e^{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}} \partial^2_{uv} C_\rho\end{aligned}$$

This implies $F(\{C_t\}_{t\geq 0})_t = \rho_t$ $\square$.

Proof of lemma 6: Write $C^\alpha_t = \alpha C_t + (1-\alpha)\tilde{C}_t$. Then, for all $(u,v) \in (0,1)^2$,

$$\begin{aligned}\rho^\alpha_t(u,v) &= F(\{\alpha C_t + (1-\alpha)\tilde{C}_t\})_t \\ &= \frac{2\pi t \partial_t C^\alpha_t - \frac{1}{2}[e^{-\Phi^{-1}(u)^2}\partial^2_{uu}C^\alpha_t + e^{-\Phi^{-1}(v)^2}\partial^2_{vv}C^\alpha_t]}{e^{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}\partial^2_{uv}C^\alpha_t} \\ &= \alpha\rho_{C_t}(u,v)\frac{\partial^2_{uv}C_t}{\partial^2_{uv}C^\alpha_t} + (1-\alpha)\rho_{\tilde{C}_t}(u,v)\frac{\partial^2_{uv}\tilde{C}_t}{\partial^2_{uv}C^\alpha_t}\end{aligned}$$

(with obvious notations) and thus:

$$|\rho^\alpha_t(u,v)| \leq |\rho_{C_t}(u,v)| \vee |\rho_{\tilde{C}_t}(u,v)|. \left[\alpha\frac{\partial^2_{uv}C_t}{\partial^2_{uv}C^\alpha_t} + (1-\alpha)\frac{\partial^2_{uv}\tilde{C}_t}{\partial^2_{uv}C^\alpha_t}\right] = 1$$

Moreover, as $\rho^\alpha_t = f(u,v)\rho_{C_t}(u,v) + (1-f(u,v))\rho_{\tilde{C}_t}(u,v)$ (where $f(u,v) = \alpha.\frac{\partial^2_{uv}C_t}{\partial^2_{uv}C^\alpha_t}$), and

$$\begin{aligned}1-\rho^\alpha_t &= f(u,v)(1-\rho_{C_t}(u,v)) + (1-f(u,v))(1-\rho_{\tilde{C}_t}(u,v)) \\ &\geq (\inf_{t,(u,v)}(1-\rho_{C_t}(u,v))) \wedge (\inf_{t,(u,v)}(1-\rho_{\tilde{C}_t}(u,v))) > 0\end{aligned}$$

because both ρ_{C_t} and $\rho_{\tilde{C}_t}$ are bounded away from 1. And likewise, $\inf_{t,(u,v)} 1+\rho^\alpha_t > 0$, which prove that ρ^α is bounded away of ± 1. Finally, it is obvious that $\rho^\alpha_t = \rho$ if t is small enough, and that it is continuous. $\square$

In order to prove that a given copula C can be attained, it is useful to know under what conditions a time-dependent mixture between a copula family in Cop and a constant copula C remains in Cop.

Lemma 10 *Let $\{C_t\}$ in Cop(ρ) for some $\rho \in (-1,1)$ and C a copula such that $\rho_C \in Corr_{BM}$*

Define $t > 0$,

$$\delta_t = \inf_{\substack{(u,v)\in(0,1)^2 \\ s.t.\, C_t(u,v)\neq C(u,v)}} \left\{ \frac{e^{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}}{2\pi|C_t(u,v)-C(u,v)|}[(1-|\rho_t(u,v)|)\partial^2_{uv}C_t \wedge (1-|\rho_C(u,v)|)\partial^2_{uv}C] \right\}$$

and suppose that $\delta_t > 0$ *for all* $t > 0$. *Let* α_t *be a function that is continuously differentiable on* $\boldsymbol{R}_+$ *and takes values in* $[0,1]$, *which is moreover constant equal to 1 on a non empty interval* $[0,\varepsilon]$, *such that, for all* $t > 0$,

$$\inf_t(\delta_t - t|\partial_t\alpha_t|) > 0$$

then the copula family $\hat{C}_t = \alpha_t C_t + (1-\alpha_t)C$ *is in* $Cop(\rho)$.
In the case where the copula family $\{C_t\}$ *is constant, equal to the Gaussian copula* C_ρ *for some* $\rho \in (-1,1)$, *the above condition reads* $\sup_t t|\partial_t\alpha_t| < \delta_C(\rho)$, *where*

$$\delta_C(\rho) = \inf_{\substack{(u,v)\in(0,1)^2 \\ s.t.\, C_t(u,v)\neq C(u,v)}} \left\{ \frac{e^{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}}{2\pi|C_\rho(u,v)-C(u,v)|}[(1-|\rho|)\partial^2_{uv}C_\rho(u,v) \wedge (1-|\rho_C(u,v)|)\partial^2_{uv}C] \right\}.$$

Proof It is trivial that $\hat{C}_t$ is indeed a copula with everywhere positive density. With obvious notations:

$$\hat{\rho}_t = \frac{2\pi t\dot{\alpha}_t(C_t - C)}{e^{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}\partial^2_{uv}\hat{C}_t(u,v)} + \alpha_t\rho_t(u,v)\frac{\partial^2_{uv}C_t}{\partial^2_{uv}\hat{C}_t}(u,v) + (1-\alpha_t)\rho_C(u,v)\frac{\partial^2_{uv}C}{\partial^2_{uv}\hat{C}_t}(u,v)$$

Thus, $\hat{\rho}_t$ is continuous, and equals ρ for t small enough.

$$|\hat{\rho}_t| \leq \frac{2\pi t|\dot{\alpha}_t||C_t - C|}{e^{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}\partial^2_{uv}\hat{C}_t(u,v)} + \alpha_t|\rho_t(u,v)|\frac{\partial^2_{uv}C_t}{\partial^2_{uv}\hat{C}_t} + (1-\alpha_t)|\rho_C(u,v)|\frac{\partial^2_{uv}C}{\partial^2_{uv}\hat{C}_t}$$

A sufficient condition for the r.h.s. to be less than 1 is, for (u,v) s.t. $C_t(u,v) \neq C(u,v)$

$$\begin{aligned} t|\dot{\alpha}_t| &\leq \frac{e^{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}}{2\pi|C_t - C|}(\partial^2_{uv}\hat{C}_t(u,v) - \alpha_t|\rho_t(u,v)|\partial^2_{uv}C_t - (1-\alpha_t)|\rho_C(u,v)|\partial^2_{uv}C_t) \\ &= \frac{e^{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}}{2\pi|C_t - C|}(\alpha_t(1-|\rho_t(u,v)|)\partial^2_{uv}C_t + (1-\alpha_t)(1-|\rho_C(u,v))|\partial^2_{uv}C) \end{aligned}$$

The expression between parenthesis in the r.h.s. is higher than $(1-|\rho_t(u,v)|)\partial^2_{uv}C_t \wedge (1-|\rho_C(u,v)|)\partial^2_{uv}C$ hence the sufficient condition.
$\delta_t > 0$ is also a sufficient condition for $\hat{\rho}_t$ to be bounded away from ± 1. For instance:

$$\hat{\rho}_t + 1 = \alpha_t\frac{\partial^2_{uv}C_t}{\partial^2_{uv}\hat{C}_t}(1+\rho_t) + (1-\alpha_t)\frac{\partial^2_{uv}C}{\partial^2_{uv}\hat{C}_t}(1+\rho_C) + \frac{2\pi t\dot{\alpha}_t(C_t - C)}{e^{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}\partial^2_{uv}\hat{C}_t(u,v)}$$

So,

$$\hat{\rho}_t + 1 \geq \left\{ \alpha_t \frac{\partial^2_{uv} C_t}{\partial^2_{uv} \hat{C}_t}(1+\rho_t) \right\} \wedge \left\{ (1-\alpha_t) \frac{\partial^2_{uv} C}{\partial^2_{uv} \hat{C}_t}(1+\rho_C) \right\} - \frac{2\pi t |\dot{\alpha}_t||C_t - C|}{e^{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}} \partial^2_{uv}\hat{C}_t(u,v)} \tag{3.22}$$

The r.h.s. is strictly positive iff, on $\{(u,v)\ s.t.\ C_t \neq C\}$,

$$\frac{2\pi t |\dot{\alpha}_t||C_t - C|}{e^{-\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}} \partial^2_{uv}\hat{C}_t(u,v)} < \left\{ \alpha_t \frac{\partial^2_{uv} C_t}{\partial^2_{uv} \hat{C}_t}(1+\rho_t) \right\} \wedge \left\{ (1-\alpha_t) \frac{\partial^2_{uv} C}{\partial^2_{uv} \hat{C}_t}(1+\rho_C) \right\}$$

i.e.

$$t|\dot{\alpha}_t| < \frac{e^{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}}{2\pi|C_t - C|} (\alpha_t \partial^2_{uv} C_t(1+\rho_t)) \wedge ((1-\alpha_t)\partial^2_{uv} C(1+\tilde{\rho}_t))$$

However, $(1+\rho_t) \geq (1-|\rho_t|)$ and a sufficient condition for the r.h.s. of (3.22) to be positive is

$$t|\dot{\alpha}_t| < \frac{e^{\frac{\Phi^{-1}(u)^2+\Phi^{-1}(v)^2}{2}}}{2\pi|C_t - C|} (\partial^2_{uv} C_t(1-|\rho_t|)) \wedge (\partial^2_{uv} C(1-|\tilde{\rho}_t|))$$

that is, $t|\dot{\alpha}_t| < \delta_t$. As $\inf_t(\delta_t - t|\dot{\alpha}_t|) > 0$, the r.h.s. of (3.22) is not only positive but $\inf_t(\hat{\rho}_t + 1) > 0$. This is true also for $\inf_t(\hat{\rho}_t - 1)$ and thus $\hat{\rho}_t$ is bounded away from ± 1.□

Before proving the proposition 11, we need the following technical lemma:

Lemma 11 *Let $0 < \varepsilon < \eta$. Consider the function*

$$\alpha_t^{\varepsilon,\eta} = \mathbb{1}_{t \leq \varepsilon} + \frac{e}{2} e^{-\frac{t-\epsilon}{\eta - t}} \mathbb{1}_{t \geq \frac{\varepsilon+\eta}{2}} + \left(1 - \frac{e}{2} e^{-\frac{\eta - t}{t-\epsilon}}\right) \mathbb{1}_{\varepsilon < t < \frac{\varepsilon+\eta}{2}}$$

α_t is continuously differentiable on $\mathbf{R}_+$, equals 1 if $t \leq \varepsilon$, 0 if $t \geq \eta$, is decreasing on $[\varepsilon, \eta]$, and

$$\forall t \geq 0,\ t|\partial \alpha_t^{\varepsilon,\eta}| \leq \frac{2\eta}{\eta - \varepsilon}$$

Proof:

$$\partial_t \alpha_t^{\varepsilon,\eta} = -\frac{e}{2} \frac{\eta - \varepsilon}{(\eta - t)^2} e^{-\frac{t-\epsilon}{\eta - t}} \mathbb{1}_{t \geq \frac{\varepsilon+\eta}{2}} - \frac{e}{2} \frac{\eta - \varepsilon}{(t - \varepsilon)^2} e^{-\frac{\eta - t}{t-\varepsilon}} \mathbb{1}_{t \leq \frac{\varepsilon+\eta}{2}}$$

Using the easy fact that the function $\frac{e^{-\frac{K}{x}}}{x^2}$, $K > 0$, $x \geq 0$ reaches its maximum at $x = \frac{K}{2}$, and the maximum thus equals $\frac{4}{e^2 K^2}$, and writing that $e^{-\frac{t-\varepsilon}{\eta - t}} = e.e^{-\frac{\eta - \varepsilon}{\eta - t}}$ and $e^{-\frac{\eta - t}{t-\varepsilon}} = e.e^{-\frac{\eta - \varepsilon}{t-\varepsilon}}$, we have,

$$|\partial_t \alpha_t^{\varepsilon,\eta}| \leq \frac{2}{(\eta - \varepsilon)} \mathbb{1}_{t \geq \frac{\varepsilon+\eta}{2}} + \frac{2}{(\eta - \varepsilon)} \mathbb{1}_{t \leq \frac{\varepsilon+\eta}{2}}$$

And thus $t|\partial_t \alpha_t^{\varepsilon,\eta}| \leq \frac{2\eta}{\eta - \varepsilon}$. □

Proof of proposition 11: With the same notations as in lemma 11, define $\alpha_t^{\varepsilon,K} = \alpha_t^{\varepsilon,K\varepsilon}$, for some constant $K > 1$ to be determined later. Consider the copulas $C_t^\varepsilon = \alpha_t^\varepsilon C_\rho + (1 - \alpha_t^\varepsilon) C$.

We want to show that the corresponding correlation is bounded by 1. And indeed, applying lemma 11, we know that

$$t|\partial_t \alpha_t^{\varepsilon,K}| \leq \frac{2K}{K-1}$$

Thus, if $\delta_C(\rho) > 2$ for some ρ, there exists a $K_0 > 1$ such that, for all ε,

$$t|\partial_t \alpha_t^{\varepsilon,K_0}| < \delta_C(\rho)$$

By virtue of lemma 10, this implies that $\{C_t^\varepsilon\} \in \mathrm{Cop}(\rho)$. $\square$

Proof of proposition 12: W.l.o.g. assume $\theta \geq 0$ and thus $\rho_\theta \geq 0$. Moreover we need the

Lemma 12 *For all $u \in (0,1)$, $u.(1-u).e^{\Phi^{-1}(u)^2/2} \leq \frac{1}{2}$.*

Proof: proving $u.(1-u).e^{\Phi^{-1}(u)^2/2} \leq \frac{1}{2}$ for all $u \in (0,1)$ is equivalent to prove that for all $x \in \mathbf{R}$, $(1-\Phi(x))\Phi(x)e^{\frac{x^2}{2}} \leq \frac{1}{2}$.

$$\begin{aligned} e^{\frac{x^2}{2}}\Phi(x) &= e^{x^2/2} \int_{-\infty}^{x} e^{-u^2/2} \frac{du}{\sqrt{2\pi}} = \int_{-\infty}^{0} e^{-u^2/2-xu} \frac{du}{\sqrt{2\pi}} \leq \int_{-\infty}^{0} e^{-u^2/2} \frac{du}{\sqrt{2\pi}}, \text{ if } x \leq 0 \\ &= \frac{1}{2} \end{aligned}$$

and thus $(1-\Phi(x))\Phi(x)e^{\frac{x^2}{2}} \leq \frac{1}{2}$ when $x \leq 0$. When $x > 0$, we have in the same manner:

$$\begin{aligned} (1-\Phi(x))e^{\frac{x^2}{2}} &= e^{x^2/2} \int_{x}^{+\infty} e^{-u^2/2} \frac{du}{\sqrt{2\pi}} = \int_{0}^{+\infty} e^{-u^2/2-xu} \frac{du}{\sqrt{2\pi}} \leq \int_{0}^{+\infty} e^{-u^2/2} \frac{du}{\sqrt{2\pi}}, \text{ if } x \geq 0 \\ &= \frac{1}{2} \end{aligned}$$

and thus $(1-\Phi(x))\Phi(x)e^{\frac{x^2}{2}} \leq \frac{1}{2}$ for all x. $\square$
Back to the proof of the proposition, recall that $\rho_\theta(u,v) = a(u,v) + a(v,u)$, with

$$a(u,v) = \theta.\frac{e^{(\Phi^{-1}(v)^2-\Phi^{-1}(u)^2)/2}v(1-v)}{1+\theta(1-2u)(1-2v)}$$

$a(u,v) \leq \theta$ for all $(u,v) \in (0,1)^2$: this is equivalent to showing that

$$e^{(\Phi^{-1}(v)^2-\Phi^{-1}(u)^2)/2}v(1-v) \leq 1+\theta(1-2u)(1-2v)$$

The r.h.s. is greater than $1-\theta$. Using the lemma 12, the l.h.s. is less than $1/2$. Eventually, $1-\theta \geq 1/2$ as $\theta \leq 1/2$. So $\rho_\theta(u,v) = a(u,v) + a(v,u) \leq 2\theta \leq 1$. $\square$

Proposition 17 *Let $\rho(t,u,v)$ be a Borelian function on $\mathbf{R}_+ \times (0,1)^2$, bounded by 1. Let $f(t,x,y) = \rho(t,\Phi(x/\sqrt{t}),\Phi(y/\sqrt{t}))$ and $g(t,x,y) = \sqrt{1-f^2(t,x,y)}$. Assume that f and g satisfy: for all $n \in \mathbf{N}^*$, $||(x_1,x_2)|| \leq n$, $||(y_1,y_2)|| \leq n$, for almost every $t \geq 0$, $||f(t,x_1,y_1) - f(t,x_2,y_2)|| \leq K_t(n)||(x_1-x_2,y_1-y_2)||$ (and the same for g) where $K_t(n)$ is finite and satisfy*

$$\int_0^T K_t^2(n)dt < +\infty, \text{ for all } T > 0$$

In other words, assume f and g are t-almost everywhere locally Lipschitz in the space variable, and that for all balls, the corresponding time-dependent Lipschitz constant is locally square integrable. Then the coupling SDE (3.10) has a unique strong solution.

This type of result is classic and dates back to Itō, see Krylov and Rozovskii [11] theorem 3.1, pp. 1254-1255 and references therein.

3.7 Formulas

This section gathers the formulas of the second order derivatives that intervene in the copula PDE for various copula families.

3.7.1 Gaussian copula

$$C_\rho(u,v) = \Phi_\rho\Big(\Phi^{-1}(u),\Phi^{-1}(v)\Big)$$

where Φ_ρ is the cdf of the bivariate normal distribution with correlation ρ, namely:

$$\Phi_\rho(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho uv)}dudv$$

The useful derivatives are:

$$\begin{cases} \partial^2_{uu}C_\rho(u,v) = \frac{-\rho}{\sqrt{1-\rho^2}}\exp\Big(-\frac{1}{2(1-\rho^2)}[(2\rho^2-1)\Phi^{-1}(u)^2+\Phi^{-1}(v)^2-2\rho\Phi^{-1}(u)\Phi^{-1}(v)]\Big) \\ \partial^2_{vv}C_\rho(u,v) = \partial^2_{uu}C_\rho(v,u) \\ \partial^2_{uv}C_\rho(u,v) = \frac{1}{\sqrt{1-\rho^2}}\exp\Big(-\frac{1}{2(1-\rho^2)}\big(\Phi^{-1}(u)^2+\Phi^{-1}(v)^2-2\rho\Phi^{-1}(u)\Phi^{-1}(v)\big)+\frac{1}{2}\big(\Phi^{-1}(u)^2+\Phi^{-1}(v)^2\big)\Big) \\ \partial_\rho C_\rho(u,v) = \frac{1}{\sqrt{1-\rho^2}}\frac{1}{2\pi}\exp\Big(-\frac{1}{2(1-\rho^2)}[\Phi^{-1}(u)^2+\Phi^{-1}(v)^2-2\rho\Phi^{-1}(u)\Phi^{-1}(v)]\Big) \end{cases}$$

A proof of the formula for $\partial_\rho C_\rho$ can be found in Plackett [17], p. 353. Alternatively, it can be

directly recovered by applying the copula PDE of two Brownian motions with deterministic correlation.

3.7.2 Student Copula

The bivariate Student copula with correlation parameter ρ and degree of freedom $\nu \in \mathbf{R}_+^*$, is defined as

$$C^{\rho,\nu}(u,v) = t^{\rho,\nu}(t_\nu^{-1}(u), t_\nu^{-1}(v))$$

where t_ν is the univariate cdf of the Student distribution:

$$t_\nu(x) = \int_{-\infty}^{x} \frac{\Gamma(((\nu+1)/2)}{\Gamma(\nu/2)} \frac{(1+\frac{w^2}{\nu})^{-\frac{\nu+1}{2}}}{\sqrt{\nu\pi}} dw$$

and $t^{\rho,\nu}$ is the cdf of the bivariate Student distribution with correlation ρ and dof ν:

$$t^{\rho,\nu}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \frac{\Gamma(((\nu+2)/2)}{\Gamma(\nu/2)} \frac{\left(1+\frac{w^2+z^2-2\rho.w.z}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}}{\nu\pi\sqrt{1-\rho^2}} dw$$

The derivatives are:

$$\begin{cases} \partial^2_{uu}C^{\rho,\nu}(u,v) & = -\frac{\Gamma((\nu+2)/2)\Gamma(\nu/2)}{\Gamma^2((\nu+1)/2)}\left(\rho + \frac{t_\nu^{-1}(u)t_\nu^{-1}(v)}{\nu}\right)(1-\rho^2)^{\frac{\nu+1}{2}}\nu^{\frac{\nu+2}{2}}\left(1+\frac{t_\nu^{-1}(u)^2}{\nu}\right)^{\nu} \\ & \quad .(t_\nu^{-1}(u)^2 + t_\nu^{-1}(v)^2 - 2\rho t_\nu^{-1}(u)t_\nu^{-1}(v) + \nu(1-\rho^2))^{-\frac{\nu+2}{2}} \\ \partial^2_{vv}C^{\rho,\nu}(u,v) & = \partial^2_{uu}C^{\rho,\nu}(v,u) \\ \partial^2_{uv}C^{\rho,\nu}(u,v) & = \frac{\Gamma((\nu+2)/2)\Gamma(\nu/2)}{\Gamma^2((\nu+1)/2)}(1-\rho^2)^{\frac{\nu+1}{2}}\nu^{\frac{\nu+2}{2}}\left(1+\frac{t_\nu^{-1}(u)^2}{\nu}\right)^{\frac{\nu+1}{2}}\left(1+\frac{t_\nu^{-1}(v)^2}{\nu}\right)^{\frac{\nu+1}{2}} \\ & \quad .\left(t_\nu^{-1}(u)^2 + t_\nu^{-1}(v)^2 - 2\rho t_\nu^{-1}(u)t_\nu^{-1}(v) + \nu(1-\rho^2)\right)^{-\frac{\nu+2}{2}} \\ \partial_\rho C(u,v) & = \frac{1}{2\pi}(1-\rho^2)^{(\nu-1)/2}\nu^{(\nu+1)/2}((1-\rho^2)\nu + t_\nu^{-1}(u)^2 + t_\nu^{-1}(v)^2 - 2\rho t_\nu^{-1}(u)t_\nu^{-1}(v))^{-\nu/2},\ \nu > 1 \end{cases}$$

Therefore, the stationary Brownian correlation function of the Student copula is

$$\begin{aligned} corr^{\rho,\nu}(u,v) = \frac{\rho + \frac{t_\nu^{-1}(u)t_\nu^{-1}(v)}{\nu}}{2}\Bigg(& e^{\frac{\Phi^{-1}(v)^2-\Phi^{-1}(u)^2}{2}}\left(1+\frac{t_\nu^{-1}(u)^2}{\nu}\right)^{\frac{\nu-1}{2}}\left(1+\frac{t_\nu^{-1}(v)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\ & + e^{\frac{\Phi^{-1}(u)^2-\Phi^{-1}(v)^2}{2}}\left(1+\frac{t_\nu^{-1}(v)^2}{\nu}\right)^{\frac{\nu-1}{2}}\left(1+\frac{t_\nu^{-1}(u)^2}{\nu}\right)^{-\frac{\nu+1}{2}}\Bigg) \end{aligned} \tag{3.23}$$

3.7.3 Archimedean copulas

The archimedean copulas are a class of copulas that takes the form $C_\varphi(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$ where φ is a continuous, strictly decreasing function from $[0,1]$ to $[0,+\infty]$ such that $\varphi(1) = 0$, where $\varphi^{[-1]}$ is the pseudo-inverse of φ defined as $\varphi^{[-1]}(t) = \mathbb{1}_{0\leq t\leq\varphi(0)}\varphi^{-1}$. φ is called the *generator* of the copula.

Clayton copula: For $\theta \geq -1$, $\theta \neq 0$, the generator of the Clayton copula with parameter θ is $\frac{1}{\theta}(t^{-\theta}-1)$ and the copula reads $C_\theta(u,v) = (u^{-\theta}+v^{-\theta}-1)_+^{-1/\theta}$. Furthermore, when $\theta > 0$, the copula admits a density and

$$\begin{cases} \partial^2_{uu}C_\theta(u,v) &= -(\theta+1)(u^{-\theta}+v^{-\theta}-1)^{-1/\theta-2}u^{-\theta-2}(v^{-\theta}-1) \\ \partial^2_{vv}C_\theta(u,v) &= \partial^2_{uu}C_\theta(v,u) \\ \partial^2_{uv}C_\theta(u,v) &= (\theta+1)(u^{-\theta}+v^{-\theta}-1)^{-2-1/\theta}u^{-\theta-1}v^{-\theta-1} \\ \partial_\theta C_\theta(u,v) &= C_\theta(u,v)\Big(-\frac{1}{\theta}\log(C_\theta(u,v)) + \frac{1}{\theta}\frac{\log(u)u^{-\theta}+\log(v)v^{-\theta}}{u^{-\theta}+v^{-\theta}-1}\Big) \end{cases}$$

Gumbel copula: This copula has generator $(-\log(t))^\theta$, for $\theta \geq 1$, and

$$\begin{cases} C_\theta(u,v) &= \exp\left(-A^{\frac{1}{\theta}}\right),\ \textit{where } A(u,v) = (-\log(u))^\theta + (-\log(v))^\theta \\ \partial^2_{uu}C_\theta(u,v) &= \frac{C_\theta(u,v)}{u^2}A^{1/\theta-2}(-\log(u))^{2\theta-2}\left[A^{1/\theta} - A(-\log(u))^{-\theta}(\theta-1-\log(u)) + \theta - 1\right] \\ \partial^2_{vv}C_\theta(u,v) &= \partial^2_{uu}C_\theta(v,u) \\ \partial^2_{uv}C_\theta(u,v) &= \frac{C_\theta(u,v)}{uv}(\log(u)\log(v))^{\theta-1}A^{1/\theta-2}\Big[A^{1/\theta}+\theta-1\Big] \end{cases}$$

Frank copula: for $\theta \neq 0$, its generator is $-\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$, and

$$\begin{cases} C_\theta(u,v) &= -\frac{1}{\theta}\log\left((1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right) \\ \partial^2_{uu}C_\theta(u,v) &= \theta.\frac{e^{-\theta u}(e^{-\theta v}-1)(e^{-\theta v}-e^{-\theta})}{(e^{-\theta}-1+(e^{-\theta u}-1)(e^{-\theta v}-1))^2} \\ \partial^2_{vv}C_\theta(u,v) &= \partial^2_{uu}C_\theta(v,u) \\ \partial^2_{uv}C_\theta(u,v) &= \theta.(1-e^{-\theta})\frac{e^{-\theta u}e^{-\theta v}}{(e^{-\theta}-1+(e^{-\theta u}-1)(e^{-\theta v}-1))^2} \end{cases}$$

Gumbel-Barnett copula: for $\theta \in (0,1]$, its generator is $\log(1-\theta\log(t))$, and

$$\begin{cases} C_\theta(u,v) & = uve^{-\theta\log(u)\log(v)} \\ \partial^2_{uu}C_\theta(u,v) & = -\theta e^{-\theta\log(u)\log(v)}(1-\theta\log(v))\log(v)\frac{v}{u} \\ \partial^2_{vv}C_\theta(u,v) & = \partial^2_{uu}C_\theta(v,u) \\ \partial^2_{uv}C_\theta(u,v) & = e^{-\theta\log(u)\log(v)}(1-\theta-\theta\log(uv)+\theta^2\log(u)\log(v)) \end{cases}$$

Archimedean copula 4.2.10 in [15], p. 116. In order to support our intuition that copulas without upper or lower tail dependence are suitable to couple Brownian, we chose one such archimedean copula, whose generator is $\log(2t^\theta-1)$, $\theta\in(0,1]$. Then,

$$\begin{cases} C_\theta(u,v) & = \frac{uv}{(1+(1-u^\theta)(1-v^\theta))^{\frac{1}{\theta}}} \\ \partial^2_{uu}C_\theta(u,v) & = \frac{(\partial_u C)^2}{C} + C\left(-\frac{1}{u^2} + \frac{u^{\theta-1}(1-v^\theta)}{1+(1-u^\theta)(1-v^\theta)}\{\frac{\theta-1}{u} + \theta\frac{u^{\theta-1}(1-v^\theta)}{1+(1-u^\theta)(1-v^\theta)}\}\right) \\ \partial^2_{vv}C_\theta(u,v) & = \partial^2_{uu}C_\theta(v,u) \\ \partial^2_{uv}C_\theta(u,v) & = \frac{\partial_u C\partial_v C}{C} - \theta\frac{u^{\theta-1}v^{\theta-1}}{1+(1-u^\theta)(1-v^\theta)}C\{1-\frac{(1-u^\theta)(1-v^\theta)}{1+(1-u^\theta)(1-v^\theta)}\} \end{cases}$$

3.7.4 FGM copulas

$C_\theta(u,v) = uv+\theta uv(1-u)(1-v)$, $|\theta|\le 1$. Obviously, $C_0=\Pi$.

$$\begin{cases} \partial^2_{uu}C_\theta(u,v) & = -2\theta v(1-v) \\ \partial^2_{vv}C_\theta(u,v) & = \partial^2_{uu}C_\theta(v,u) \\ \partial^2_{uv}C_\theta(u,v) & = 1+\theta(1-2u)(1-2v) \end{cases}$$

3.7.5 Plackett copula

The Plackett copula is $C_\theta(u,v) = \frac{1}{2(\theta-1)}((1+(\theta-1)(u+v))-\sqrt{(1+(\theta-1)(u+v))^2-4uv\theta(\theta-1)})$, $\theta>0$, and $C_1(u,v)=uv$.

$$\begin{cases} \partial^2_{uu}C_\theta(u,v) & = \frac{2\theta(\theta-1)v(v-1)}{((1+(\theta-1)(u+v))^2-4uv\theta(\theta-1))^{\frac{3}{2}}} \\ \partial^2_{vv}C_\theta(u,v) & = \partial^2_{uu}C_\theta(v,u) \\ \partial^2_{uv}C_\theta(u,v) & = \frac{\theta(1+(\theta-1)(u+v-2uv))}{((1+(\theta-1)(u+v))^2-4uv\theta(\theta-1))^{\frac{3}{2}}} \end{cases}$$

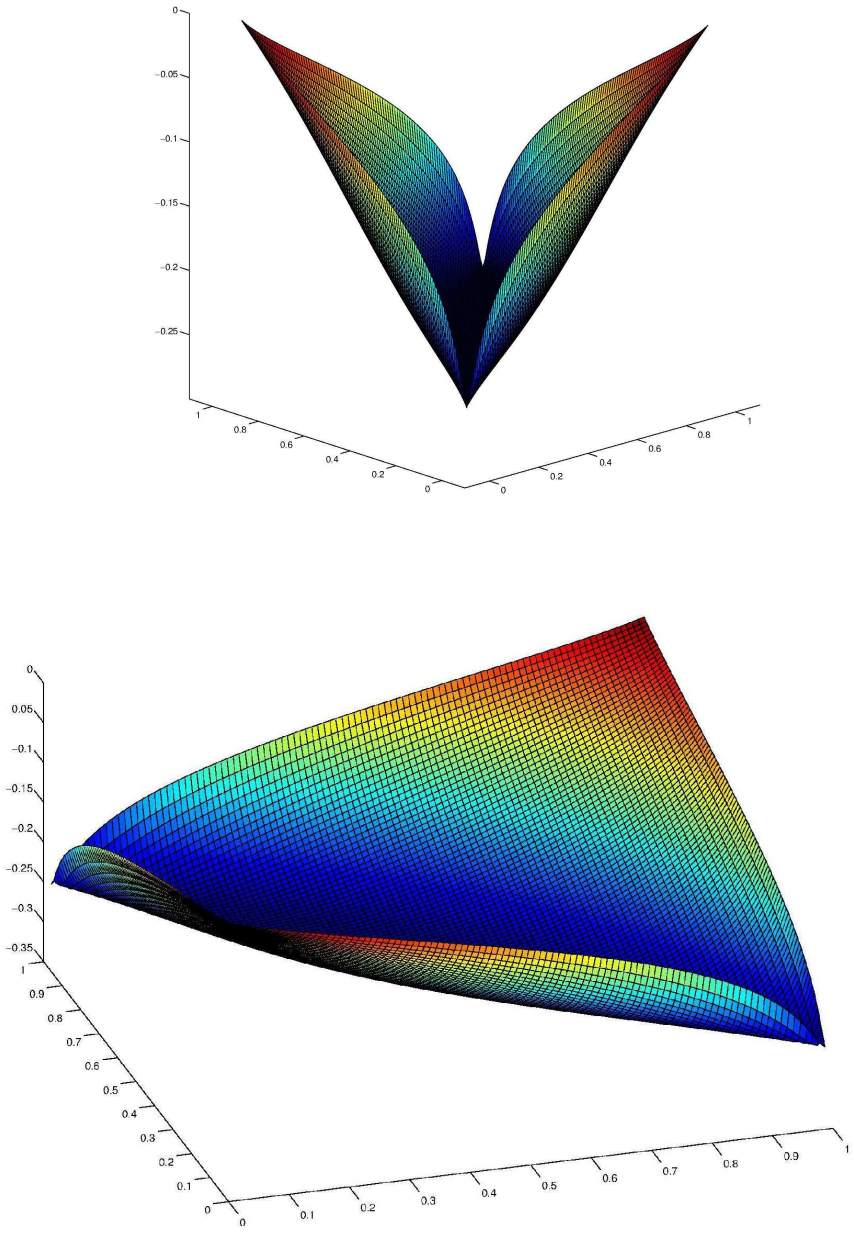

Figure 3.2: FGM Copula stationary correlation. Above: $\theta = -1$. Below: $\theta = 1$.

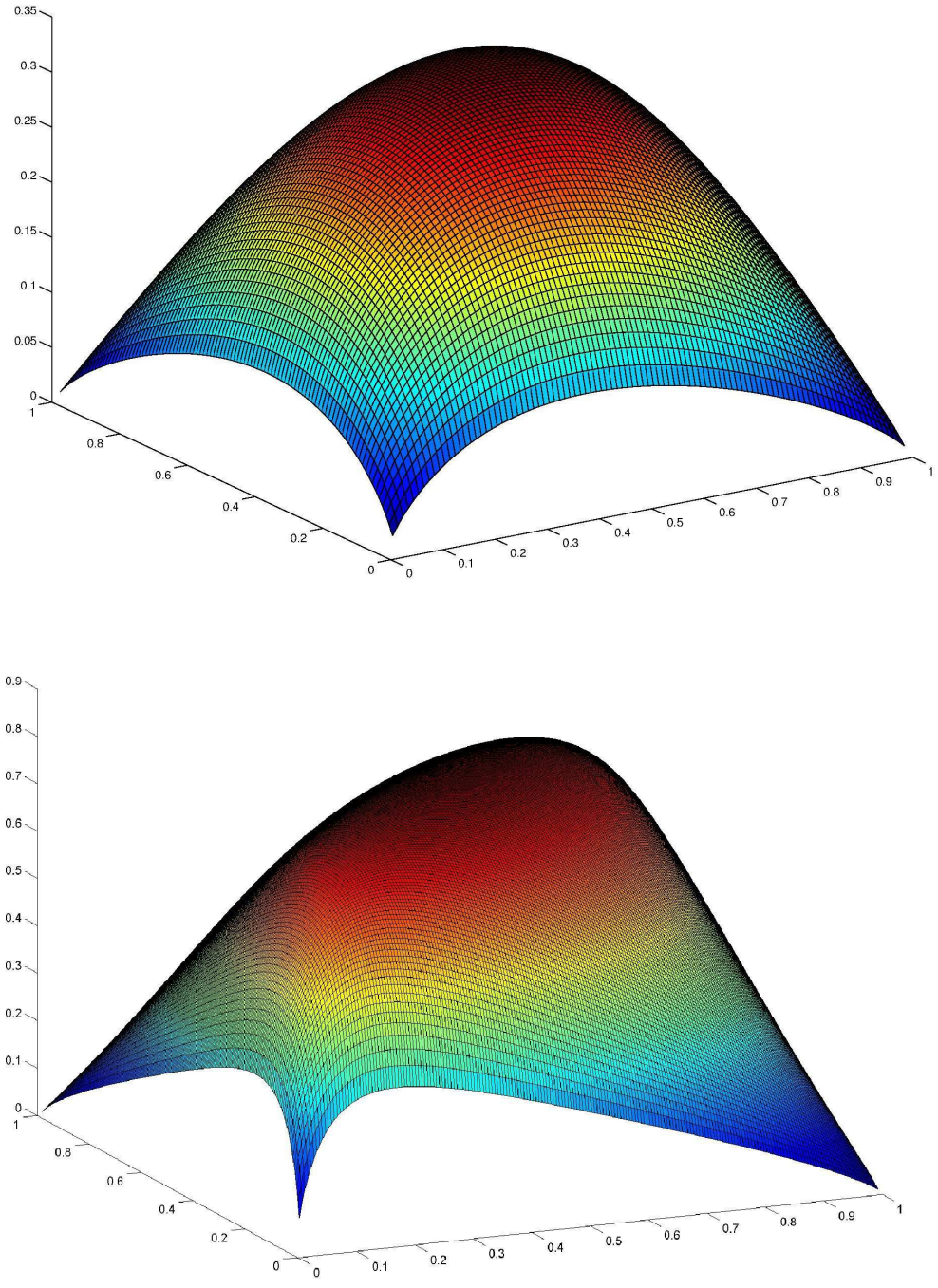

Figure 3.3: Plackett Copula stationary correlation. Above: $\theta = 2$. Below: $\theta = 10$.

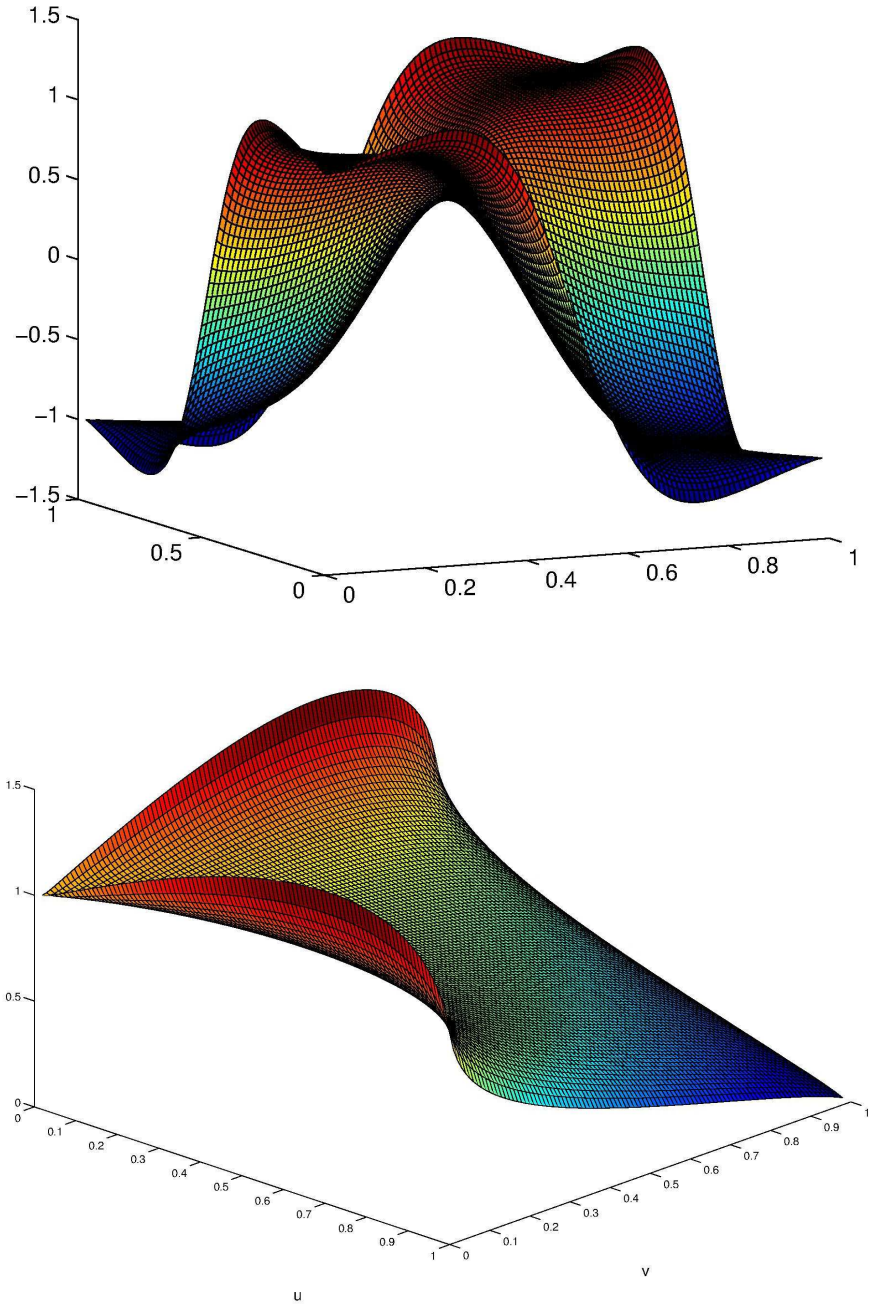

Figure 3.4: Explosive behavior of some copulas stationary correlations. Above: Student stationary correlation function, $\rho = 0.5$, $\nu = 0.4$. Below: Clayton correlation function, $\theta = 2$.

Conclusion

This thesis studied two aspects of dependence modeling. The first one is the understanding and modeling of the multivariate dependence, i.e. the dependence between random vectors. The second aspect is the dependence between continuous-time stochastic processes, and more precisely modeling the dependence between continuous stochastic processes with the help of copulas.

Optimal transport theory provides a means to generalize the notions of quantiles and comonotonicity to the multivariate setting.Therefore it has been used to define multivariate risk measures. We implemented and studied a method that computes an approximation of the optimal transport map when the initial measure and the target measure are continuous ; This method was compared to several classical algorithms and proved to behave efficiently. However, several questions remain. The convergence speed and complexity of the quasi-Newton algorithm, respectively $O(N^{5/2})$ and $O(1/\sqrt{N})$ for the transport problem in $\mathbf{R}^2$, remain to be proved theoretically. Moreover, the IPFP produces high numerical errors on the boundary of the support of the initial measure; it seems that this is not the only algorithm that exhibits this kind of behavior, and it would be interesting to test the performance of such algorithms on distributions with periodic support to avoid the difficulties that arises on the boundary of the support. Eventually, we mentioned others algorithms, and there remains to do a complete comparison across more existing algorithms.

The second part of the thesis proposed a definition of extreme dependence between fixed multivariate laws of probability. This definition is based on the notion of covariogram, defined as the set of all possible cross-covariance matrix between the multivariate marginals. We give a method that is numerically tractable to compute extreme couplings; this procedure can be used to define trajectories of couplings that starts at some coupling whose cross-covariance matches a given cross-covariance matrix and goes to an extreme coupling. This trajectory can be used to stress the dependence between the multivariate marginals, for instance in problems of portfolio allocation or in the risk management of options on several underlyings. Furthermore, the parameterization of these trajectories allows to define an index of the strength of the dependence between the marginals. However, this index is not invariant by transforms of the marginals and the question remains to determine whether a measure of multivariate

dependence can actually be derived from this index. Extreme couplings can be also derived by maximization of cross-covariance matrices with respect to some conic orders. While the variational characterization of the extreme couplings makes it clear that these couplings are a particular case of extreme couplings, the relation between the various notions of extreme dependence associated to conic orders deserves to be further investigated, as well as the relevance of using a given conic order in practical applications.

Finally, the third chapter gives some answers as to the spatial dependence that can be attained by two univariate Markovian diffusions. More specifically, the coupling of two Brownian motions by stationary copulas is highlighted (that is Brownian motions with a constant spatial copula after some time). We provided case-by-case results showing that some copulas were admissible to model such dependence, while others (including Student, Clayton, Gumbel copulas) were not. However, deriving sufficient and necessary conditions that can be used in practice to determine which copulas are attainable by coupled Brownian motions (or admissible to model stationary dependence between Brownian motions) is still an open question. We treated the bivariate case, providing an integrated form of the Kolmogorov forward PDE that describes the evolution of the spatial copula of coupled Markovian diffusions. The multivariate case is more complex to tackle, as on top of necessary boundedness of the correlation coefficients, the correlation matrix needs also to be nonnegative, which complicate further the characterization of multivariate copula that are admissible to couple several diffusions. Note also that we focused on a particular coupling problem: the marginal diffusions are Markovian. One could also consider the case where the bivariate diffusion is Markovian but not the marginals (i.e. the drifts and volatilities depend on the state of both marginals). In this case, the integration of the Kolmogorov forward equation can not be made as in the case we studied, and the link between the copula family $\{C_t\}_t$ and the correlation function (or correlation matrix) is less clear.

Finally, a subject of potentially high interest is the application of optimal transport techniques to diffusion equations. For instance a Markov functional model describing a strong dependence between two multivariate diffusions X_t and Y_t could be $\tilde{Y}_t = \nabla\varphi_t(X_t)$, where $\nabla\varphi_t$ is the optimal transport map between the law of X_t and the law of Y_t. $\hat{Y}_t$ is a multivariate process with the same one dimensional marginals as Y_t, which means that $\tilde{Y}_t \sim Y_t$ for all t. Such a model raises the question of the smoothness of the maps $\nabla\varphi_t$ both in space and time, and of the possibility of sampling trajectories from such models within a sensible amount of time and with an acceptable accuracy.

Index