ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

EXTREME DEPENDENCE FOR MULTIVARIATE DATA

Damien BOSC
Alfred GALICHON

December 2010

Cahier n° 2010-35

DEPARTEMENT D'ECONOMIE
Route de Saclay
91128 PALAISEAU CEDEX
(33) 1 69333033
http://www.enseignement.polytechnique.fr/economie/
mailto:chantal.poujouly@polytechnique.edu

EXTREME DEPENDENCE FOR MULTIVARIATE DATA

DAMIEN BOSC[†] ALFRED GALICHON[§]

Abstract. We present a generalized notion of extreme multivariate dependence between two random vectors which relies on the extremality of the cross-covariance matrix between these two vectors. Using a partial ordering on the cross-covariance matrices, we also generalize the notion of positive upper dependence. We then quantify the strength of dependence between two given multivariate series using an entropic distance to extremally dependent distributions. We apply this method to build idnices of exposure to a financial environment, and to do stress-tests on the correlation between two sets of financial variables.

JEL codes: C58, C02.

Date: First version is March 19, 2010. The present version is of December 24, 2010. The authors thank Rama Cont for a question which was the starting point of this paper and Guillaume Carlier and Alexander Sokol for helpful conversation. Galichon gratefully acknowledges support from Chaire Axa "Assurance et Risques Majeurs," Chaire FDIR "Socially Responsible Investments and Value Creation," Chaire EDF-Calyon "Finance and Développement Durable," and FiME, Laboratoire de Finance des Marchés de l'Energie (www.fime-lab.org).

[†] *Economics Department, École polytechnique, and AXA-IM. E-mail: damien.bosc@polytechnique.edu.*

[§] *Economics Department, École polytechnique. E-mail: alfred.galichon@polytechnique.edu.*

Introduction

Extreme dependence, and the closely related notion of comonotonicity are important concepts in various fields. In economics, it is central in the economics of insurance (following the seminal work of Borch [4] and Arrow [1], [2]), in economic theory (see [26], [14] and [23]), and in mathematical statistics (see [8], [21], [19], [28]).

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:

$$\mathbf{E}(XY) = \sup_{\substack{\tilde{X} \sim X \\ \tilde{Y} \sim Y}} \mathbf{E}(\tilde{X}\tilde{Y}), \tag{0.1}$$

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 paper 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. Fist, 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. [25]) would be to consider the following optimization problem

$$\max_{\substack{\tilde{X}\sim X\\ \tilde{Y}\sim Y}} \mathbf{E}(\tilde{X}\cdot\tilde{Y}) \tag{0.2}$$

where $\cdot$ is the scalar product in $\mathbb{R}^d$. This program is a multivariate extension of the covariance maximization problem (0.1) and to define as extreme the distribution of the pair $(\tilde{X},\tilde{Y})$ solution to the above problem. However this is not fully satisfactory, as it does not take into account the cross-dependence between X_i and Y_j for $i\neq j$. This generalization seems therefore somehow arbitrary.

A more satisfactory generalization is based on the idea that both positive and negative extreme dependence are obtained by the maximization of a non-zero 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 (0.2) where one replaces the scalar product by any non-zero 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 non-zero 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 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 yields to 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. Eventually, 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 foreign risk exposure as well as a prospective application to progressive stress tests of dependence.

The paper is organized as follows: notations and definition are first recalled. 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 coupling with 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 Appendix B.

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 (two) real valued random variables: this is the subject of the theory of copulas. In most of this article 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 of the form $\mathbf{E}(X_i Y_j)$ 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 $\mathbf{M}_{I,J}(\mathbb{R})$, we define by $M \cdot N$ their matrix scalar product $M \cdot N = Tr\,(M'N)$. 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. 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.

1. Multivariate extreme dependence

In this section we describe our proposed notion of multivariate extreme dependence. Introduce the covariogram as the set of cross covariance matrix $\mathbf{E}_\pi(X_i Y_j)$ in $\mathbf{M}_{I,J}(\mathbb{R})$ for any $\pi \in \Pi(p, q)$. More formally:

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 [25]) the covariogram is itself a convex compact subset of $\mathbf{M}_{I,J}(\mathbb{R})$.

Figure 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 $I = J = 2$. Here 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 cross-covariance matrix in the convex hull of this curve, with the coupling of independence projecting 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 $\left(\mathbf{E}_\pi(X_i Y_j)\right)_{ij}$ lies on the boundary of the covariogram $\mathcal{F}(p, q)$.*

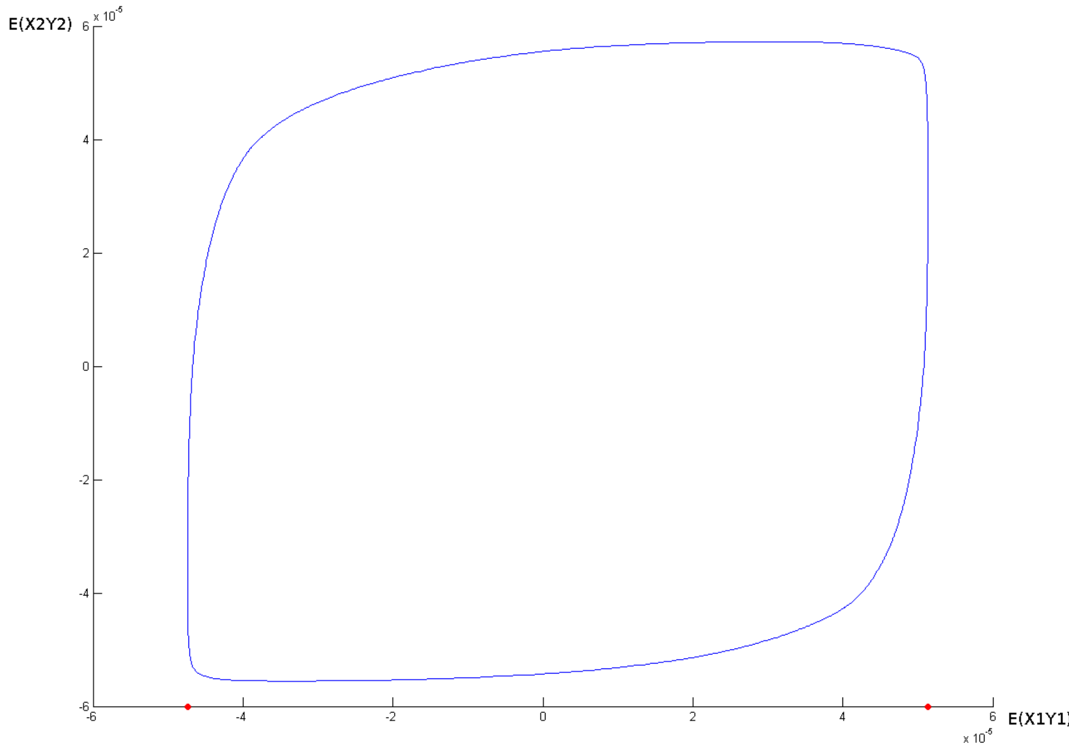


FIGURE 1. Example of a 2 dimensional section of a covariogram

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

$$Tr\left(M'\sigma_{X,Y}\right) = \mathbf{E}(X'MY), \text{ for all } M \in \mathbf{M}_{I,J}(\mathbb{R}) \tag{1.1}$$

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

Theorem 1. *The following conditions are equivalent:*

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

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

$$Tr\left(M'\sigma_{\pi}\right) = \sup_{\pi' \in \Pi(p,q)} Tr\left(M'\sigma_{\pi'}\right)$$

or equivalently

$$\mathbf{E}_{\pi}(X'MY) = \sup_{\pi' \in \Pi(p,q)} \mathbf{E}_{\pi'}(X'MY); \tag{1.2}$$

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 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 anticomonotic otherwise.

2. Positive extreme dependence

The aim of this section is to propose a generalization of Fréchet copula of upper dependence in 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 $C_\pi\,(x, y) = \min(C_1(x_1, \ldots, x_I), C_2(x_1, ..., x_J))$. But according to the so-called 'Impossibility theorem' (see [17]), C_π is a copula function if and only if C_1 and C_2 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 are considered 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 (appendix A 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:

$$K(C) = \{y \in \mathbf{M}_{I,J}(\mathbb{R}) | x \cdot y \geq 0, \, \forall x \in C\} \tag{2.1}$$

Considering cones of this form only 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 $\mathbf{M}_{I,J}(\mathbb{R})$

$$A \succ_{K(C)} B \; iff \; A - B \in Int(K(C))$$

Note that $Int(K(C)$ is merely $\{y \in \mathbf{M}_{I,J}(\mathbb{R}) | x \cdot y > 0, \, \forall x \in C\}$.
Let $K = K(C)$ be a cone as above.

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 $\succ_K$ generated by K will be said to have* positive extreme dependence *with respect to order $\succ_K$.*

The following results fully characterize maximal couplings in terms of maximal correlation.

Theorem 2. *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

$$Tr\left(M'\sigma_\pi\right) = \sup_{\pi' \in \Pi(p,q)} Tr\left(M'\sigma_{\pi'}\right)$$

or equivalently

$$\mathbf{E}_\pi(X'MY) = \sup_{\pi' \in \Pi(p,q)} \mathbf{E}_{\pi'}(X'MY); \tag{2.2}$$

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. Evidently, this result is a close parallel to Theorem 1 except that M is constrained to belong to C. As a consequence the *positive extreme couplings* are a particular case of the *extreme couplings*.

Once again the interpretation in dimension 1 is straightforward: X and Y have positive extreme dependence (for the order in $\mathbf{R}$) iff they are comonotonic.

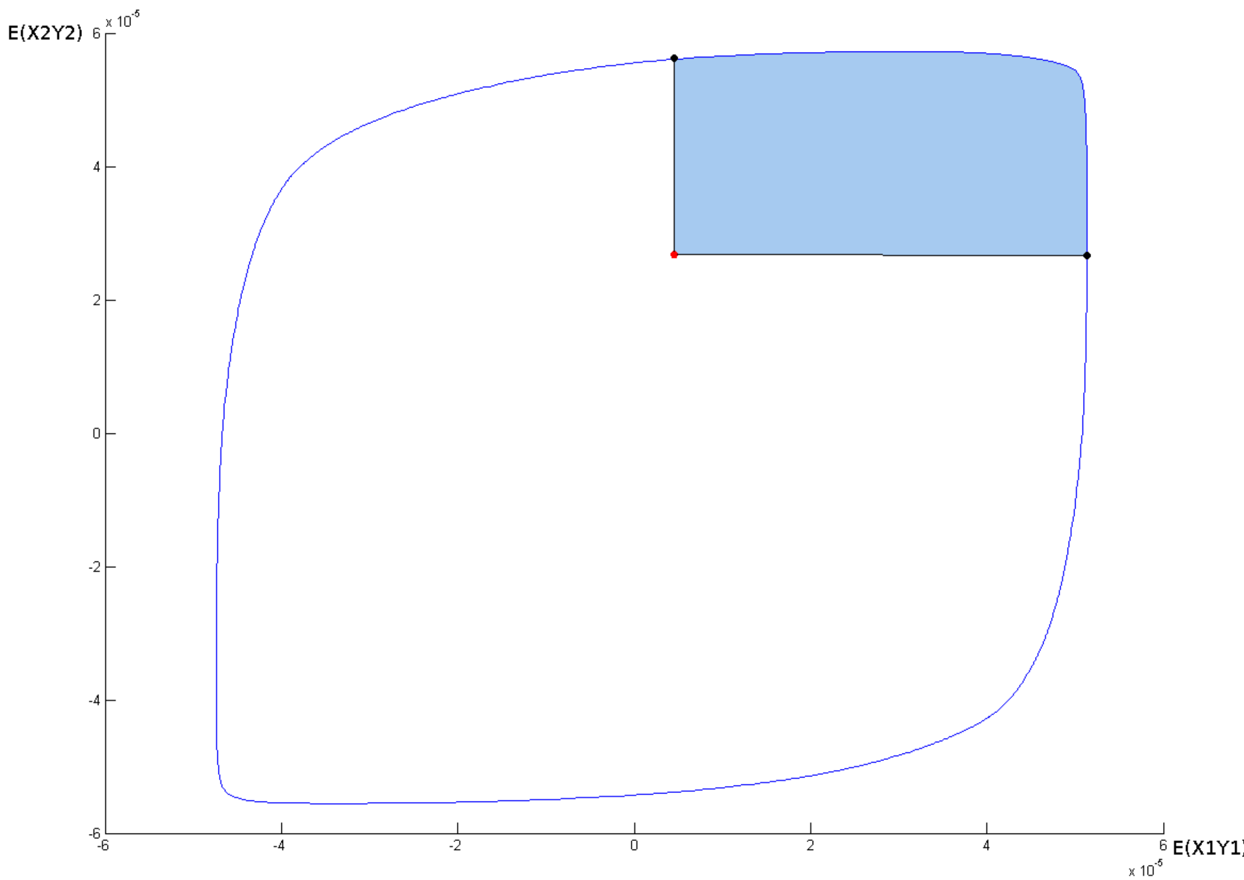


FIGURE 2. In blue: location of the couplings dominating the red coupling

To better understand the relation between those two types of coupling, let us go back to the two dimensional section of the covariogram discussed in the previous section, and take

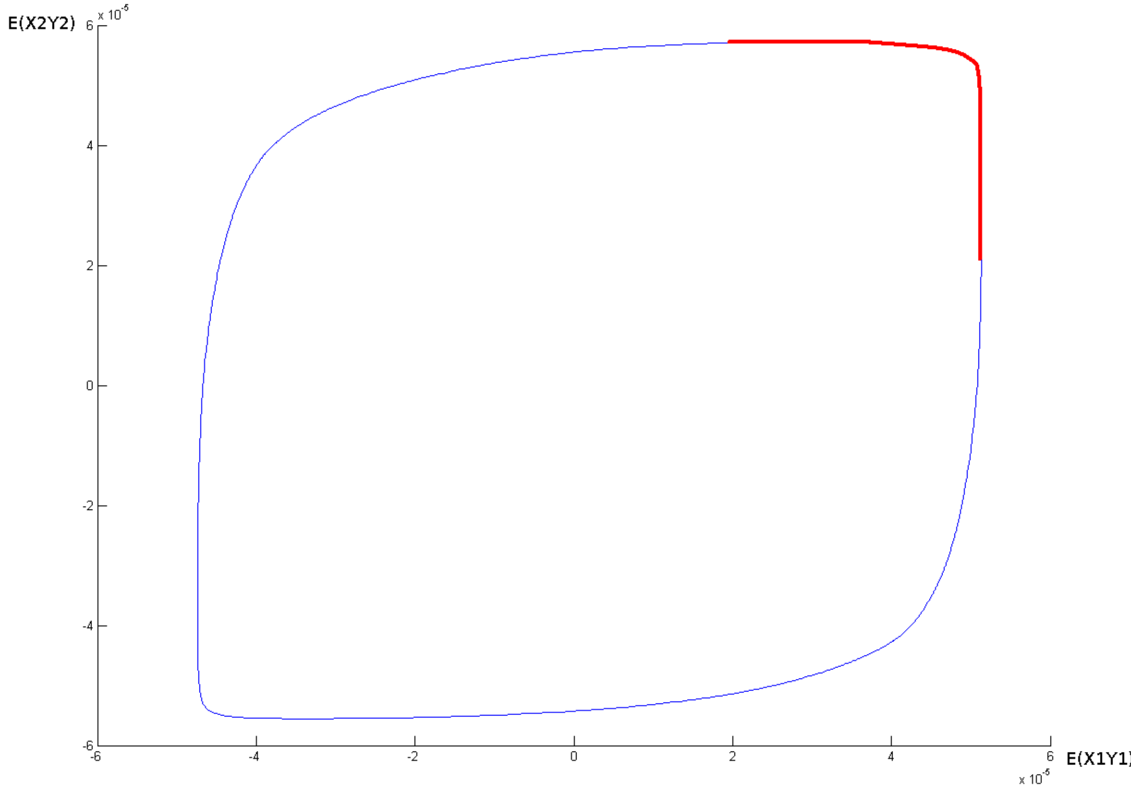


FIGURE 3. Maximal couplings on the boundary

for K the positive orthant of $\mathbb{R}^2 \times \mathbb{R}^2$. The blue region in Figure 2 is the set of couplings dominating the red coupling 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 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 3. They are on the red curve, in the upper right corner of the covariogram, and forms only a 'small' part of the couplings of extreme dependence.

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

Example 1 (Orthant order). *Let $M^{+}_{I,J}(\mathbb{R})$ (resp. $M^{++}_{I,J}(\mathbb{R})$) denotes the set of 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 all its coefficients positive: 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 $C = \{S \in S^{+}_{n}(\mathbb{R}) | Tr(M) = 1, \}$ 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 example shows that this ordering allows various positive extreme couplings. A first remark is that the maximum correlation coupling is positive extreme, by taking $M = Id$ in theorem 2.

A trivial example is to 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 if $X_1 = Y_1$. However it satisfies (2.2) with $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and can be qualified as a maximal coupling.

Example 3 (Hermitian order). *The previous example might seem restrictive as it does use a transform of the cross-covariance matrix one can consider as arbitrary. However defining* *an order on $M_n(\mathbb{R})$ would allow for the direct use of the cross-covariance matrix. 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 $\hat{M} \in M_n(\mathbb{C})$ *of M by setting*

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

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

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

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

Finally, we can state that $\sigma_{X,Y}$ is maximal if and only if there exists a non null matrix M with nonnegative hermitian transform such that X and MY have the maximum correlation.

3. 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, the projection of $\hat{\pi}$ will never be *exactly* located on the boundary of the covariogram, although it can be very near from it. Our problem is then to *associate an extreme coupling with $\hat{\pi}$.*

We will then modify (1.2) so that its solutions project on inner points of the covariogram. This is done by a classic method in physics called entropic relaxation or entropy penalization.

3.1. **Entropic relaxation.** We introduce heat in (1.2) by means of an entropy term; it becomes

$$W(M, T) := \max_{\pi \in \Pi(p,q)} \mathbf{E}(X'MY) + TEnt(\pi) \tag{3.1}$$

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 (3.1).

A first remark is that W is homogeneous in (M,T): for all $\lambda \in \mathbb{R}$, $W(\lambda M, \lambda T) = W(M,T)$. Hence T is fixed at 1 and we shall write $W(M)$ and π_M instead of $W(M,T)$ and $\pi_{M,T}$.

Our aim is to find a matrix M such that $\hat{\pi}$ and π_M have the same location in the covariogram; in other words they have the same cross-covariance matrix: $\sigma_{\hat{\pi}} = \sigma_{\pi_M}$.

The gradient of W is given by the envelope theorem : $\nabla_M W = \sigma_{\pi_M}$. This remark implies that M is the solution of the following variational problem

$$\min_{M \in M_n(\mathbb{R})} W(M) - \sigma_{\hat{\pi}} \cdot M \tag{3.2}$$

W is a convex function as a supremum of affine functions in M and consequently the objective function in (3.2) is convex as well: this is a classic unconstrained convex minimization problem. A similar approach in a matching context is used in [12].

Figure 4 shows the projection of π_M for a great number of randomly chosen matrices M. The red dot is the projection of $\hat{\pi}$. One sees that any inner point of the covariogram can be attained by a properly chosen π_M.

3.2. **Numerical solution.** It can be shown that the optimal π_M in (3.1) obeys a Schrödinger equation (see appendix B.3):

$$\log \pi_M(x,y) \quad = x'My + u(x) + v(y), \quad u \in L^1(dp), v \in L^1(dq)$$

In other words, the log-likelihood of π_M 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 π_M has the right marginals p and q. This is the purpose of the long 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 [22] 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.

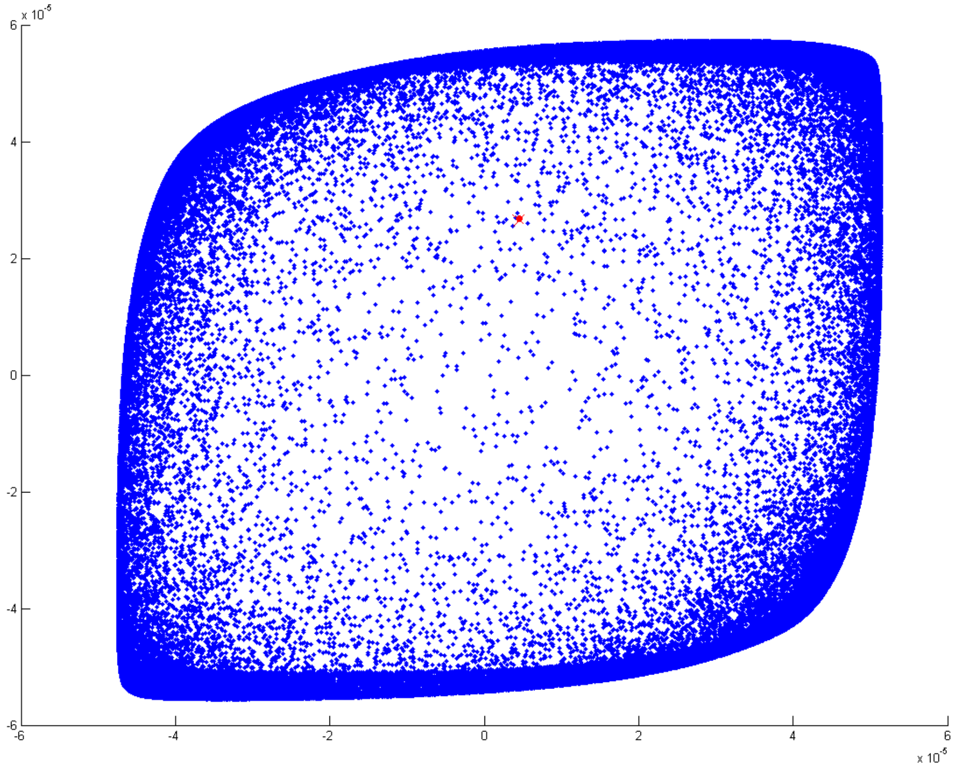


FIGURE 4. Projection of various π_M

π_n has the following form:

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

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 simply 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

$$\begin{cases} e^{u_{n+1}(x)} & = & \frac{p(x)}{\int e^{x'My+v_n(y)}dy} \\ e^{v_{n+1}(y)} & = & \frac{q(y)}{\int e^{x'My+u_{n+1}(x)}dx} \end{cases}$$

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

$$\begin{cases} \int e^{x'My+u(x)+v(y)}dy & = & p(x) \\ \int e^{x'My+u(x)+v(y)}dx & = & q(y) \end{cases}$$

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 [22] for more details).

An important remark is that in the case of discrete distributions p and q, the previous formulae simplify greatly: the problematic computation of the integral in the recursion which yields the explosion of the complexity in the case of continuous distribution becomes simple linear relations.

Indeed, if $r(x, y) = \frac{e^{x'My}}{\sum_{x,y} e^{x'My}}$, the recursion becomes

$$\begin{cases} e^{v_{n+1}(y)} & = & \frac{q(y)}{\sum_x r(x,y)e^{u_n(x)}} \\ e^{u_{n+1}(y)} & = & \frac{p(x)}{\sum_y r(x,y)e^{v_{n+1}(y)}} \end{cases}$$

Eventually the convex minimization problem can be solved by any gradient descent type algorithm. The BFGS algorithm is used in the examples below.

3.3. **Derivation of the extreme coupling.** We recall that our aim is to associate some extreme coupling, a coupling which projects onto the boundary of the covariogram, with an inner coupling (i.e a coupling which projects inside the covariogram).

The previous algorithm yields a particular matrix $\hat{M}$ and a coupling $\pi_{\hat{M}}$ such that $\sigma_{\hat{\pi}} = \sigma_{\pi_{\hat{M}}}$. 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 (3.1). 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 (3.1) 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.

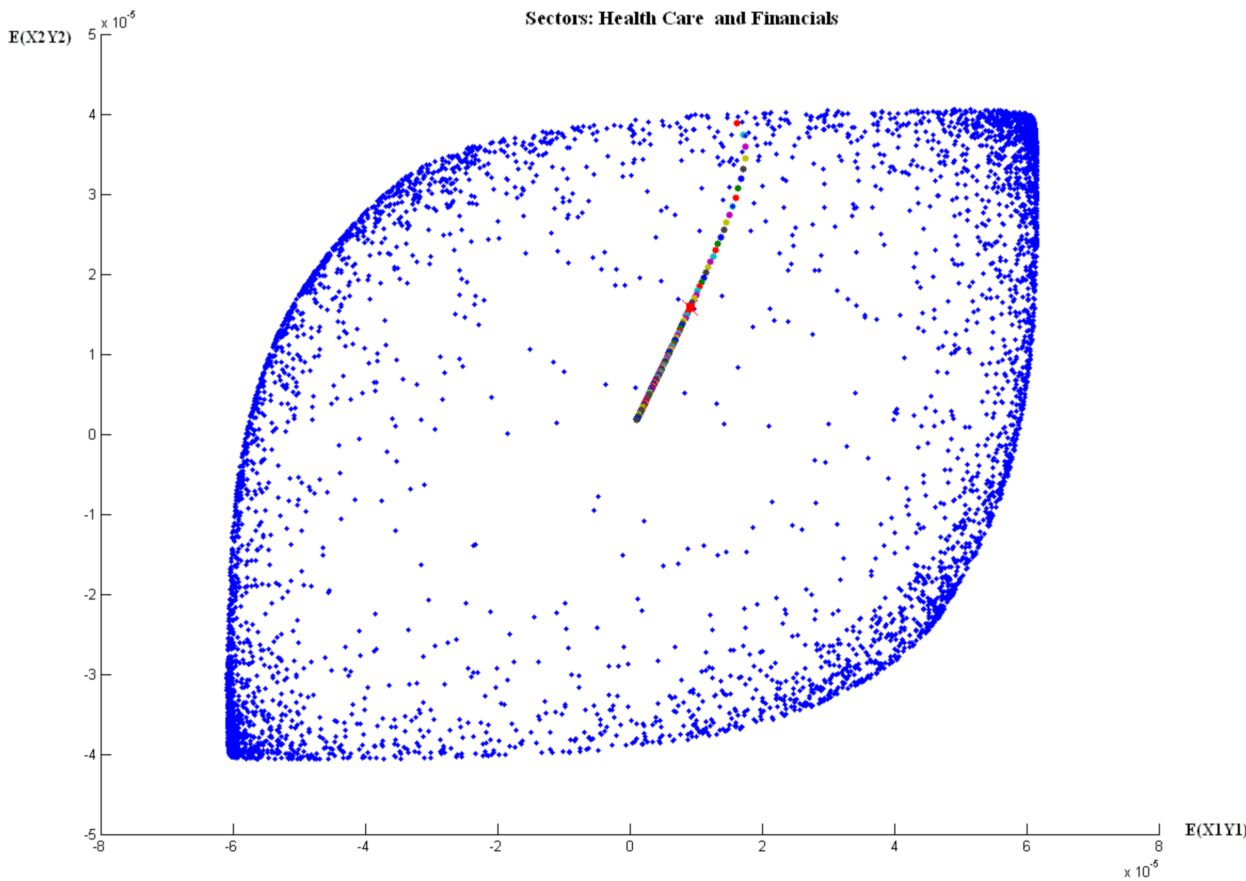


FIGURE 5. A trajectory towards an extreme coupling

Figure 5 summarizes this idea: each point on the curve is the projection of a $\pi_{\hat{M},T_n}$. At 'high' temperature, say 10, 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.

The temperature controls the strength of the dependence: fixing M, $T_1 < T_2$ yields couplings π_{M,T_1} and π_{M,T_2} with the latter being a coupling of stronger dependence than the first one.

The matrix $\hat{M}$ can be seen as an *affinity matrix*: in the limit of $T = 0$, the extreme coupling is a $\pi_{\hat{M},0}$ which 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 $\pi_{\hat{M},0}$, hence the name of affinity matrix.

4. Some applications

We apply the previous technique 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 1 gives summary statistics (the three first variables corresponds to S&P sectors, the last third to Eurostoxx).

Table 1. Summary Statistics

Mean Returns	$\begin{pmatrix} 1.03\times10^{-4} & -1.13\times10^{-4} & 1.67\times10^{-4} & 1.16\times10^{-4} & -1.37\times10^{-4} & 3.99\times10^{-4} \end{pmatrix}$
Variance	$\begin{pmatrix} 1.36\times10^{-4} & 7.65\times10^{-4} & 1.16\times10^{-4} & 1.14\times10^{-4} & 4.15\times10^{-4} & 1.12\times10^{-4} \end{pmatrix}$
Correlation matrix	$\begin{pmatrix} 1 & & & & & \\ 0.66 & 1 & & & & \\ 0.76 & 0.62 & 1 & & & \\ 0.22 & 0.10 & 0.19 & 1 & & \\ 0.26 & 0.33 & 0.25 & 0.49 & 1 & \\ 0.22 & 0.16 & 0.22 & 0.67 & 0.58 & 1.00 \end{pmatrix}$
Cross-Covariance	$\begin{pmatrix} 2.74\times10^{-5} & 3.05\times10^{-5} & 2.13\times10^{-5} \\ 6.04\times10^{-5} & 1.88\times10^{-5} & 5.52\times10^{-5} \\ 2.66\times10^{-5} & 4.62\times10^{-5} & 2.56\times10^{-5} \end{pmatrix}$

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.

The data differs noticeably from log-normality: the log-returns for each index and sector exhibits an excess kurtosis (i.e. corrected by 3) above 6 in each case.

4.1. **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 $= \frac{\lvert\lvert\sigma_M - \sigma_{\hat{\pi}}\rvert\rvert}{\lvert\lvert\sigma_{\hat{\pi}}\rvert\rvert}$	$\approx 0.1\%$	$< 0.2\%$

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.

4.2. **Financial applications.** The first application exploits further the optimal matrix $\hat{M}$. It consists in performing a singular value decomposition on it in order to deduce indices of extreme dependence. This is related to the idea of canonical correlation.

The second one is based on the idea of the trajectory towards the boundary we evoked above. This could be used in construction of stress test scenarios involving multivariate variables. Instead of picking the maximum correlation coupling as the 'worst dependence scenario', one could have a gentler approach and consider the extreme coupling $\pi_{\hat{M},0}$ computed above. It is an extreme dependence coupling more in line with the cross-covariance structure of the empirical coupling $\hat{\pi}$: we expect its projection to be located nearer from the projection of $\hat{\pi}$ than the projection of the maximum correlation coupling. In a second time the trajectory of $\pi_{M,T}$, $T \to 0$ is a mean to progressively increase the dependence using a single parameter, and could yield a continuous family of scenarios of increasing dependence.

4.3. **Index of extreme dependence.** 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 the above maximum, is the highest diagonal element of the diagonal matrix that intervenes in the singular value decomposition of the matrix $\sigma_{XX}^{-1/2}\sigma_{XY}\sigma_{YY}^{-1/2}$ ([15]).

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 with nonnegative entries. In particular

$$\mathbf{E}_{\pi_{\hat{M},0}}\left((\sqrt{S}U'X)'(\sqrt{S}V'Y)\right) = \max_{\pi\in\Pi(p,q)} \mathbf{E}_{\pi}\left((\sqrt{S}U'X)'(\sqrt{S}V'Y)\right)$$

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 $\pi_{\hat{M},0}$). Here, under the proper law of probability, $\sqrt{S}U'$ 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 (1.2). 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 $(\sqrt{S}U'X, \sqrt{S}'Y)$ 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 scale 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 two indices most correlated to one another under the law of the extreme coupling. These indices are each one composed of three assets, which are actually portfolios formed of the components of the original index: we speak of *indices of foreign risk.*

4.3.1. *Portfolios stress testing.* In order to underline the necessity of accounting properly for the multivariate dependence, the problem of one-period allocation is addressed.

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 $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_m)$. In the examples below, the assets are S&P Sector Index, 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.

Table 2. Summary Statistics

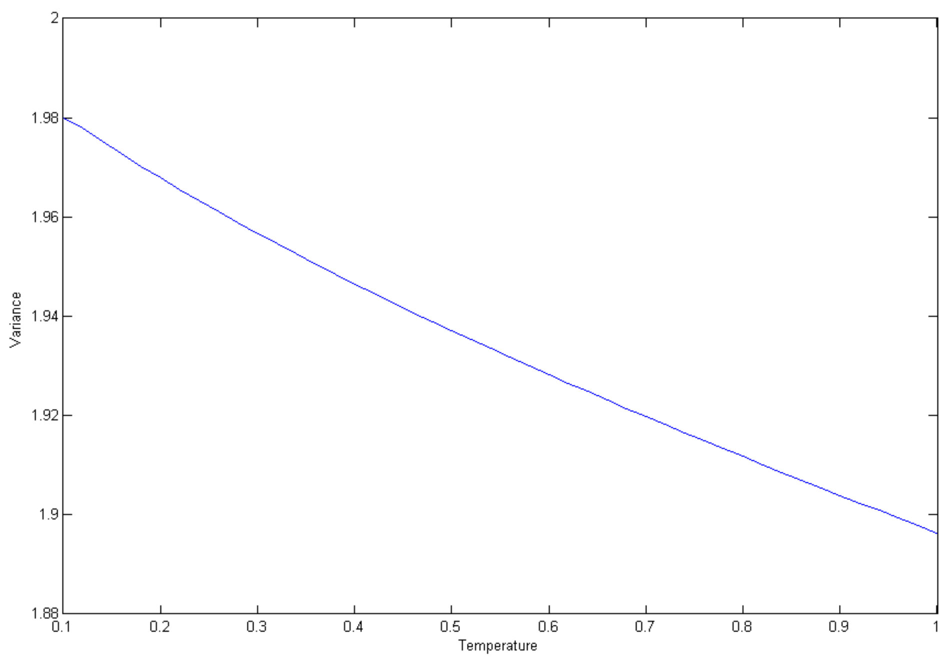
Mean Returns	$\begin{pmatrix} 2.89{\times}10^{-4} & 1.67{\times}10^{-4} & 1.03{\times}10^{-4} & -1.13{\times}10{-}4 & 1.97{\times}10^{-4} & 2.01{\times}10^{-4} & 1.85{\times}10^{-4} \end{pmatrix}$
Variance	$\begin{pmatrix} 3.59{\times}10^{-4} & 1.16{\times}10^{-4} & 1.36{\times}10^{-4} & 7.65{\times}10^{-4} & 1.92{\times}10^{-4} & 9.84{\times}10^{-5} & 3.25{\times}10^{-4} \end{pmatrix}$
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	$\begin{pmatrix} 1.35{\times}10^{-4} & 9.84{\times}10^{-5} \\ 2.44{\times}10^{-4} & 1.37{\times}10^{-4} & 3.25{\times}10^{-4} \\ 1.41{\times}10^{-4} & 9.21{\times}10^{-5} & 1.27{\times}10^{-4} & 1.16{\times}10^{-4} \\ 1.52{\times}10^{-4} & 9.78{\times}10^{-5} & 1.45{\times}10^{-4} & 9.48{\times}10^{-5} & 1.35{\times}10^{-4} \\ 3.62{\times}10^{-4} & 1.83{\times}10^{-4} & 3.73{\times}10^{-4} & 1.85{\times}10^{-4} & 2.11{\times}10^{-4} & 7.65{\times}10^{-4} \\ 1.85{\times}10^{-4} & 1.05{\times}10^{-4} & 1.50{\times}10^{-4} & 1.08{\times}10^{-4} & 1.19{\times}10^{-4} & 2.14{\times}10^{-4} & 1.92{\times}10^{-4} \end{pmatrix}$

Correlation is higher than in the above examples as the sectors are industrial sectors on a single index, the S&P500.

The sensitivity of the risk of portfolio allocation with respect to the dependence between X and Y is highlighted by monitoring the risk that arises when the multivariate dependence is mismatched by an investor.

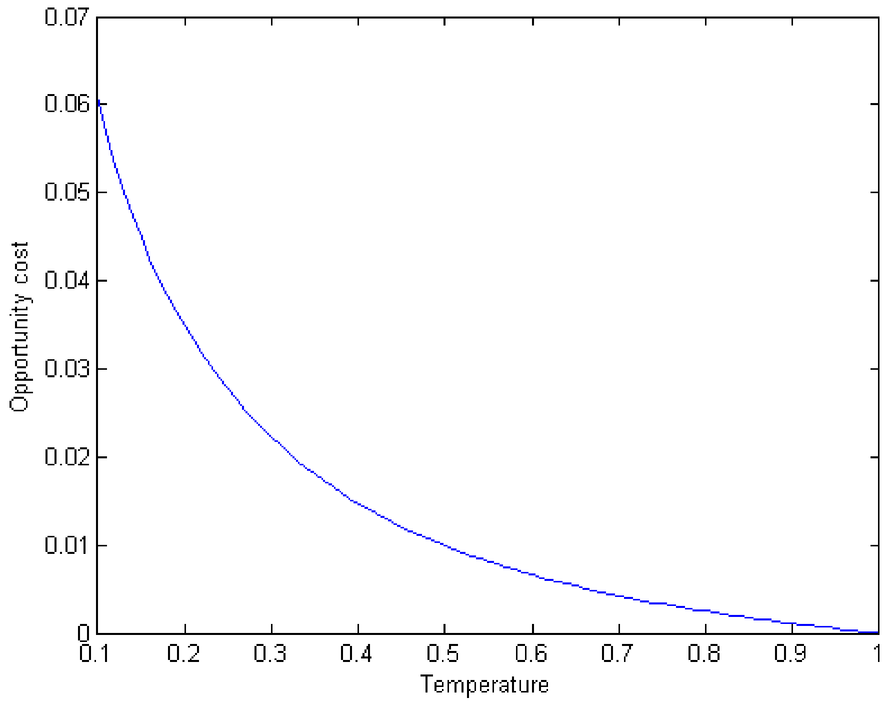
This latter solves a classic Markowitz allocation problem, with an investment horizon of one year: $\max_{\sum_i w_i = 1} \mu \cdot w - \frac{\lambda}{2} w' \Sigma w$. μ are the expected yearly returns of the stocks and Σ the covariance matrix of the returns. We assume that both μ and Σ are the standard 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 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. Only the dependence between X and Y is stressed tested: the covariance between X and Y only is modified, while their means remain unchanged.

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 to 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 is increased, compared to the initial covariance matrix Σ. In a first time, 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 first graph below shows this effect. It represents the variance $w'\Sigma_T w$ as a function of the temperature.



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 allocation do not change. 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 dimensions 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 mean to stress test 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 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 to 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 in this case 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 to generate 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.

4.3.2. *Options pricing.* This method of increasing the multivariate dependence also works for rainbow options (options on several underlyings) pricing. As a case study, consider the underlyings $X_1, \dots, X_n, Y_1, \dots, Y_m$. It is assumed that they all follows log-normal diffusions, with parameters

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

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 performs well, but mitigates the gain by selecting the lowest payoff between $(\max_i X_T^i - K)_+$ and $(\max_j Y_T^j - K)_+$.

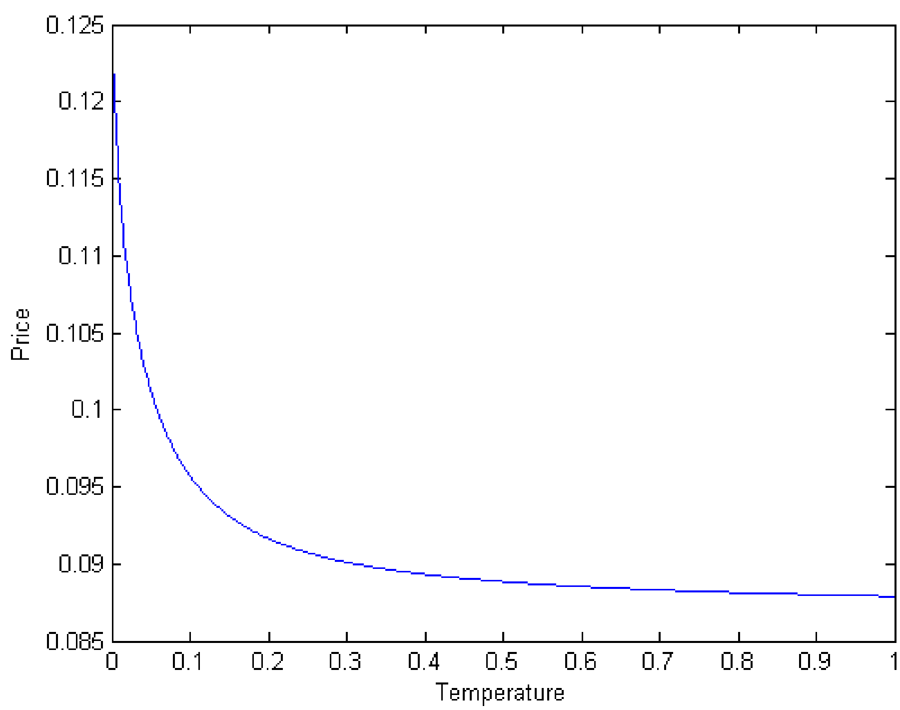
The terminal distribution of the underlyings is discretized by means of optimal quantization; the discrete marginals of vectors X and Y obtains. Their atoms are respectively denoted x_T^i and y_T^j. For each specification of a cross-covariance matrix A between X and Y, a trajectory $\pi_T(A)$ is obtained 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 follow; 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.



The above graph displays the price as a function of the temperature. It 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 minima and hence the maximum of these minima tends to be higher, which leads to 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, i.e it is still lower than the non-stressed price.

5. Conclusion

A broad complaint in applied statistics is the "curse of dimensionality": models that have a simple, computationally tractable form in dimension one become very complex, both computationaly 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.

Appendix A. 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.
This results applies of course when $\mathbf{M}_{I,J}(\mathbb{R})$ is replaced by any euclidean spaces.

Appendix B. Proof of the results in main text

B.1. **Proof of Theorem 1.**

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 non null $M \in \mathbf{M}_{I,J}(\mathbb{R})\backslash\{0\}$ 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 $M \in \left(\mathbb{R}^I \times \mathbb{R}^J\right) \setminus \{0\}$ such that

$$M \cdot \sigma_\pi = \sup_{\pi' \in \Pi(p,q)} M \cdot \sigma_{\pi'}$$

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

Equivalence between (ii) and (iii) follows from a well-known result in Optimal Transport theory, the Knott-Smith optimality criterion (see [25], Th. 2.12). ∎

B.2. **Proof of Theorem 2.**

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 1 (Variational characterization of maximality)**.**

$$\sigma_\pi \text{ maximal iff } \sup_{\pi' \in \Pi(p,q)} \inf_{M \in C} (\sigma_{\pi'} - \sigma_\pi) \cdot M = 0$$

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

Proof of proposition 1. First, note that the function

$$f : (\pi', M) \in \Pi(p,q) \times C \mapsto (\Sigma_{\pi'} - \Sigma_\pi) \cdot M$$

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

$$\max_{\pi' \in \Pi(p,q)} \min_{M \in C} f(\pi', M) = f(\bar{\pi}, \bar{M}) = \min_{M \in C} \max_{\pi' \in \Pi(p,q)} f(\pi', M) \tag{B.1}$$

This is a consequence of a classical minmax theorem by Fan [10]: a continuous function over a product of compacts convex sets embedded in normed linear spaces, 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 [25] for instance. Moreover f is linear in M and π', and continuous in both arguments. Finally, $\Pi(p,q)$ can be embedded in the space of Radon measures over $\mathbb{R}^d \times \mathbb{R}^d$ endowed with the bounded Lipschitz norm. We refer to Villani [25] chapter 7. for full 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 $\sigma_{\pi'}$ one has $\sigma_{\pi'} - \sigma_\pi \notin K_+^*$, which means that for some $M \in C$, $(\sigma_{\pi'} - \sigma_\pi) \cdot M \leq 0$, hence

$$\sup_{\Pi(p,q)} \inf_C (\sigma_{\pi'} - \sigma_\pi) \cdot M \leq 0$$

Thanks to the compacity of K, we can apply the minmax formula B.1 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 π' such that $\sigma_{\pi'} - \sigma_\pi \in K_+^*$. Thanks again to the compacity of C, $\inf_C (\sigma_{\pi'} - \sigma_\pi) \cdot M > 0$ and the reverse implication is proved. ∎

As a consequence, we are now ready to prove theorem 2.

Proof of theorem 2. Because of the previous proposition, a coupling π such as (X, MY) is an optimal transport plan shall satisfy

$$\mathbb{E}_\pi(X \cdot MY) = \sup_{\pi' \in \Pi(p,q)} \mathbb{E}_{\pi'}(X \cdot MY)$$

As $\mathbb{E}_\pi(X \cdot MY) = \sigma_\pi \cdot M$, we conclude with the proposition 1. ∎

B.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 $\mathbb{R}^n \times \mathbb{R}^n$.

$$\max_{\pi \in \Pi(p,q)} \mathbf{E}(X'MY) + TEnt(\pi) = \int x'My\pi(x,y) - T\log\pi(x,y)dxdy$$

$$= \max_{\pi \in \mathcal{M}_+^{\mathcal{E}}(\mathbb{R}^n\times\mathbb{R}^n)} \Bigg\{ \min_{\substack{\phi \in L^1(dp)\\ \psi\in L^1(dq)}} \int (x'My - T\log\pi(x,y))\pi(x,y)dxdy$$

$$- \left[\int (\phi(x)+\psi(y))d\pi(x,y) - \int \phi dp - \int \psi dq \right] \Bigg\} \quad \circledast$$

where $\mathcal{M}_+^{\mathcal{E}}(\mathbb{R}^n \times \mathbb{R}^n)$ is the set of nonnegative Radon measures on $\mathbb{R}^n \times \mathbb{R}^n$ 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 (3.1) 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 [22] 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 in this setting, which is the desired result.

References

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

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

[3] Bertsekas, D. P., "A new algorithm for the assignment problem," *Mathematical Programming* 21, pp. 152–171, 1981.

[4] K. Borch, "Equilibrium in a reinsurance market," *Econometrica* 30, pp. 424–444, 1962.

[5] Carlier, G., Dana, R.-A., Galichon, A., "Pareto efficiency for the concave order and multivariate comonotonicity," working paper, 2010.

[6] Carlier, G., Galichon, A., and Santambrogio, F., "From Knothe's transport to Brenier's map and a continuation method for optimal transport," *SIAM Journal on Mathematical Analysis*, 2010.

[7] Choo, E., and A. Siow, "Who Marries Whom and Why," *Journal of Political Economy*, 114, pp. 175–201, 2006.

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

[9] Ekeland, I., Galichon, A., Henry, M. "Comonotonic measures of multivariate risks," forthcoming in *Mathematical Finance*, 2010.

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

[11] Galichon, A., and Henry, M., "Dual theory of choice under multivariate risks," working paper, 2010.

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

[13] Hardy, G.H., Littlewood, J.E. , and Pólya, G., *Inequalities*, 2d ed., University Press, Cambridge, 1952.

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

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

[16] Nagasawa, M., *Schrödinger equation and diffusion theory*, Birkhaüser, 1993.

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

[18] Puccetti, Scarsini, "Multivariate comonotonicity," *Journal of Multivariate Analysis* 101, pp. 291–304, 2010.

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

[20] Rachev, S., and Rüschendorf, L., *Mass Transportation Problems. Volume I: Theory*, and *Volume II: Applications*, New York: Springer, 1998.

[21] Rüschendorf, L., "Fréchet-bounds and their applications," in: *Advances in Probability Distributions with Given Marginals* (Rome, 1990), Math. Appl., vol. 67, Kluwer Acad. Publ., Dordrecht, 1991, pp. 151 187.

[22] Rüschendorf, L., "Convergence of the Iterative Proportional Fitting Procedure", *The Annals of Statistics*, Vol. 23, No 4, 1160-1174, 1995.

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

[24] E. Schrödinger, "Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique," *Ann. Inst. H. Poincaré* 2, pp. 269–310, 1932.

[25] Villani, C., *Topics in Optimal Transportation*, Providence: American Mathematical Society, 2003.

[26] M.E. Yaari, "The dual theory of choice under risk," *Econometrica* 55, pp. 95–115, 1987.

[27] J. J. Kosowsky, Alan L. Yuille, "The invisible hand algorithm: Solving the assignment problem with statistical physics," *Neural Networks* 7(3), pp. 477–490, 1994.

[28] V.M. Zolotarev, "Probability metrics," *Theory Probab. Appl.* 28, pp. 278–302, 1983.