

Parameterizing correlations: a geometric interpretation

Francesco Rapisarda * Damiano Brigo Fabio Mercurio
Product and Business Development Group, Banca IMI

Introduction

Finance as a whole is linked to correlations among the prices of assets (stocks, bonds, ...) or more generally of basic financial quantities (including interest rates, credit spreads, ...). In fact, the concept of diversification itself has to do with sensibly exploiting the correlations among such quantities. It is therefore fundamental to be able to safely build and manipulate correlation structures, with which to model the joint behaviour of financial variables, both for pricing purposes and for risk management. We hint at the fact that linear correlation is a satisfactory measure of dependence only for jointly Gaussian variables, finding its correct generalization in the notion of copula function. However, when considering *instantaneous* returns, the multi-dimensional Brownian motion dynamics typically adopted in finance is Gaussian and therefore correlation parameterization maintains its interest. In this paper, whenever correlation is associated to a dynamics, it is meant as *instantaneous* correlation.

Correlations (often meant as “correlations between returns of financial variables” – be they stock prices, credit spreads, interest rates or else) are usually expressed in terms of matrices, *i.e.* tables of numbers. Matrices representing correlations between variables must satisfy four basic properties in order to really bear statistical and financial significance:

- A1) all their entries must lie in the interval $[-1, 1]$ (nothing goes beyond perfect correlation or perfect anticorrelation);
- A2) diagonal entries must be equal to one (a variable is always perfectly correlated with itself);
- A3) the matrix must be symmetric (correlation between variable a and variable b is equal to correlation between b and a);
- A4) the matrix must be *positive semidefinite*. This has to do with the fact that the variance of a portfolio, whose correlation matrix is C , is $\sigma_P^2 = \mathbf{x}^T C \mathbf{x} \geq 0$, where \mathbf{x} is the array of weights of the constituent financial variables, each multiplied by the standard deviation of the respective variable.

Notation. We denote (column) vectors by lowercase bold letters and matrices by capital letters. The i -th element of vector \mathbf{v} is denoted by $(\mathbf{v})_i$. The superscript T denotes transposition.

While the first three properties, A1) to A3), can easily be checked by inspection in a plausible

*Corresponding author

correlation matrix, the fourth is in general tougher to test, requiring explicit computation either of eigenvalues or of principal minors.

There are different ways to construct what looks like a correlation matrix for a number of assets. The most widespread, when possible, is that of computing historical correlations: if the time series of the financial variables at hand are available, one could compute estimates of the correlation matrix in standard ways, see [4]. However, a well-known problem is that if the length of the time series is smaller than the number of financial variables, positive definiteness is not guaranteed. If one considers that the basic financial entities involved in a Value at Risk (VaR) estimation for a whole portfolio count in the hundreds or even thousands, it is evident that most financial databases will not contain enough historical data to render the computation of a covariance matrix reliable.

Another possibility for building correlation matrices is through calibration of a model (in which the correlation matrix fits as a constituent input) to a set of financial data, in much the same spirit lying at the heart of the determination of the implied volatility parameter for a set of plain vanilla options. This happens for instance in the case of the Libor Market Model (LMM) where the data for a given options market (the swaptions' one) may be used to infer implied correlations for the basic rates underlying the swaptions market [2, 7, 9].

The third possibility is through sheer imagination. Suppose that a trader (or a risk manager) is called into a derivative deal involving many underlyings, trusts a model to price such deal, and wants to have an idea of how the price of the deal depends on the correlations among the underlyings' returns. What he/she will do is to start changing the entries of the correlation matrix (previously determined through historical or implied evaluation) in a fashion reflecting his/her expectation regarding correlation movements. Quite often this procedure will end up with an apparently satisfactory correlation matrix, which is however not positive definite, although the perturbation may look small to the inexperienced eye.

All this highlights the need for a parameterization of a generic correlation matrix – a parameterization guaranteeing that the matrix will fulfill all four properties above. In fact, after imposing A1) to A3), the positive semidefiniteness A4) is typically checked a posteriori by calculating, for instance, the matrix eigenvalues. A suitable parameterization for which A1) to A4) are automatically satisfied is thus fundamental for many practical applications.

In this paper we will revise the “standard angles parameterization” (SAP) procedure described for example in Pinheiro and Bates [5], recently applied to financial modeling by Rebonato and Jäckel [8], and discuss a geometric interpretation to it. This interpretation is based on viewing a correlation matrix as the result of the scalar products of a suitable set of unit vectors in a multidimensional space, each rotated from all the others by generalized Euler angles. Acting this way, some common pitfalls in the application of the SAP method are avoided, and a practical suggestion is made: exploiting the intuitive nature of this approach can lead to more efficient optimization schemes when calibrating a reduced form model to a desired correlation structure. Moreover, thanks to the new geometrical view, we will recast the SAP in a more parsimonious way, leading to the “triangular angles parameterization” (TAP), which is actually the form of SAP described in [5]. A geometric relationship of the TAP with the so called spherical parameterization that has not been pointed out in [5] is established. Finally, we consider an application to a demanding problem – that of sensibly parameterizing instantaneous correlations in the LMM.

The standard angles parameterization

In a recent paper, Rebonato and Jäckel [8] introduced to the financial world the SAP for a generic correlation matrix. Noting that a physically meaningful $n \times n$ correlation matrix C should be symmetric and positive semidefinite, and as such can be decomposed as the product

$$(1) \quad C = BB^T$$

the two authors suggest that a possible parameterization consists in choosing matrix $B = (b_{ij})$ as

$$(2) \quad b_{ij} = \begin{cases} \cos \theta_{ij} \prod_{k=1}^{j-1} \sin \theta_{ik} & \text{for } j < n \\ \prod_{k=1}^{n-1} \sin \theta_{ik} & \text{for } j = n \end{cases}$$

for $i = 1, \dots, n$, where in particular $b_{i1} = \cos \theta_{i1}$.

This parameterization helps the financial practitioner in two ways. In the first place, suppose one is interested in analyzing the effect of varying correlations on the price of an option written, say, on a basket composed of n stocks. Then, the simplest thing to do is to start changing the entries of the original matrix C , still preserving its symmetry. In most cases this will lead to a shocked matrix \tilde{C} which is not positive semidefinite anymore, thus rendering it inapplicable in any pricing engine. A piece of machinery that allows to obtain a matrix satisfying A1) to A4) and as similar as possible to the shocked correlation matrix would be of help.

A second reason for parameterizing correlations lies in the need for obtaining *reduced rank models* for a given correlation structure, in order to reduce the computational effort when pricing/hedging an instrument through simulation. This is fundamental when the number of underlying variables is considerable. Typical examples are the calculation of VaR figures for a bank portfolio or the valuation of derivatives under the LMM.

In the following we will show that the SAP amounts to applying a set of Jacobi rotations to an original unit vector lying in the n -dimensional Cartesian space, thus obtaining a set of vectors $\{\mathbf{v}_i\}$ whose scalar product $\mathbf{v}_i \cdot \mathbf{v}_j = c_{i,j}$ is the (i, j) entry of the initial correlation matrix C . For this, we first need to revise the theory of Jacobi rotations.

Jacobi rotations

The Jacobi, or Givens, rotations help in introducing zeros in a vector. Take for instance vector $\mathbf{w} \in \mathbb{R}^n$. Denoting by x_i , $i = 1, \dots, n$ the i -th coordinate in \mathbb{R}^n , a Jacobi rotation by angle θ of $\mathbf{w} = (w_1, \dots, w_n)^T$ in the (x_i, x_k) plane is given by applying the following $n \times n$ matrix G

$$(3) \quad G(i, k; \theta) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos \theta & \cdots & \sin \theta & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & -\sin \theta & \cdots & \cos \theta & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

to \mathbf{w} . Matrix G differs from the identity matrix only in entries $(i, k), (k, i), (i, i), (k, k)$.

The effect of applying G to \mathbf{w} is to perform a clockwise rotation of \mathbf{w} in the (x_i, x_k) plane of θ radians. In fact, the m -th component of $G(i, k; \theta)\mathbf{w}$ is

$$(4) \quad (G(i, k; \theta)\mathbf{w})_m = \begin{cases} w_m & \text{for } m \neq i, m \neq k \\ w_i \cos \theta + w_k \sin \theta & \text{for } m = i \\ -w_i \sin \theta + w_k \cos \theta & \text{for } m = k. \end{cases}$$

If θ is such that $\cos \theta = \frac{w_k}{\sqrt{w_i^2 + w_k^2}}$ and $\sin \theta = -\frac{w_i}{\sqrt{w_i^2 + w_k^2}}$, *i.e.* $\theta = -\arctan \frac{w_i}{w_k}$, this procedure zeroes the i -th component of the resulting vector and adjusts the k -th component so as to preserve the norm of \mathbf{w} .

Iterative application of this procedure in different planes yields a vector with only one nonzero component, say the first, $(w, 0, \dots, 0)^T$, and norm $|w|$ equal to the norm of the original vector. In particular, starting from a unit vector, we can end up with any of the unit vectors in the canonical basis of \mathbb{R}^n . Of course, the reverse is also true. We can, in fact, start from a unit vector in the canonical basis of \mathbb{R}^n and, through application of counterclockwise Jacobi rotations,¹ obtain any unit vector in \mathbb{R}^n .

The SAP in view of Jacobi angles

We denote by \mathbf{e}_i the i -th unit vector in the canonical basis of \mathbb{R}^n , namely $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ (the i -th component is 1, the others are 0).

By applying a set of consecutive counterclockwise Jacobi rotations to vector \mathbf{e}_1 by angles $\theta_{m,i}$ in the planes (x_i, x_{i+1}) , $i = 1, \dots, n-1$, we exactly reproduce the SAP of the m -th column of matrix B^T of Eq. (2). The scheme is the following:

- rotate first by $-\theta_{m1}$ in the (x_1, x_2) plane:

$$(5) \quad G(1, 2; -\theta_{m1}) \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \theta_{m1} & -\sin \theta_{m1} & 0 & \cdots & 0 \\ \sin \theta_{m1} & \cos \theta_{m1} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \theta_{m1} \\ \sin \theta_{m1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- rotate then by $-\theta_{m2}$ in the (x_2, x_3) plane:

$$(6) \quad G(2, 3; -\theta_{m2}) \begin{pmatrix} \cos \theta_{m1} \\ \sin \theta_{m1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cos \theta_{m2} & -\sin \theta_{m2} & 0 & \cdots & 0 \\ 0 & \sin \theta_{m2} & \cos \theta_{m2} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \cos \theta_{m1} \\ \sin \theta_{m1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \theta_{m1} \\ \cos \theta_{m2} \sin \theta_{m1} \\ \sin \theta_{m1} \sin \theta_{m2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

¹A counterclockwise rotation by an angle α is equivalent to a clockwise rotation by an angle $-\alpha$.

- Iterate the above procedure in the next planes, up to plane (x_{n-1}, x_n) ; this yields the SAP for the m -th column of B^T , call it \mathbf{v}_m .

Essentially then the SAP boils down to rotating the unit vector lying along the first axis by suitable sets of angles (we remind that rotations preserve the norm: if we start by a unit vector and rotate it, we will always obtain a unit vector). Every column of B^T is nothing but a rotated vector, and the rotation angles $\theta_{i,k}$ are such that the scalar product of the generic i -th and j -th columns coincides with the (i, j) entry of C .

Finding the correlation matrix that best approximates \hat{C}

We denote by $\rho_{i,j}$ the (i, j) entry of the modified matrix \hat{C} . In order to find the correlation matrix that best approximates \hat{C} , we look for rotation angles $\theta_{i,k}$ such that the scalar product of the generic i -th and j -th columns, $\mathbf{v}_i(\theta) \cdot \mathbf{v}_j(\theta)$, is as close as possible to $\rho_{i,j}$. To this end, a multidimensional nonlinear minimization of a suitable cost function must be performed. Each modeler can decide which cost function is more suitable on the basis of his/her need, intuition, or preference. The easiest cost function to think of is

$$(7) \quad \chi^2(\theta_{1,1}, \theta_{1,2}, \dots, \theta_{1,n-1}, \dots, \theta_{n,1}, \theta_{n,2}, \dots, \theta_{n,n-1}) = \sum_{i,j=1}^n (\mathbf{v}_i(\theta) \cdot \mathbf{v}_j(\theta) - \rho_{ij})^2,$$

i.e. the square of the Frobenius norm [3] of the difference $C - \hat{C}$.

A similar optimization can be performed when simply in need to find which parameters $\theta_{i,k}$ produce the original matrix C : we just have to set $\rho_{i,j} = c_{i,j}$. Being C a correlation matrix, the cost function (7) has a null minimum value, whose corresponding θ 's yield the desired parameterization.

A perplexity comes about when comparing the number of free parameters (the angles), which is equal to $n(n-1)$ in the formulation of Eq. (2), and the number of data points to be fitted (the $\rho_{i,j}$'s), which is equal to $(n-1) + (n-2) + \dots + 1 = n(n-1)/2$. There are too many free parameters (the double!), something that typically yields a degeneracy in the minima of a cost function such as that of Eq. (7).

The reason for this degeneracy can be illustrated with a simple example. Take $n = 2$ so that

$$B = \begin{pmatrix} \cos \theta_{1,1} & \sin \theta_{1,1} \\ \cos \theta_{2,1} & \sin \theta_{2,1} \end{pmatrix} \quad \text{and} \quad BB^T = \begin{pmatrix} 1 & \cos(\theta_{1,1} - \theta_{2,1}) \\ \cos(\theta_{1,1} - \theta_{2,1}) & 1 \end{pmatrix}$$

It is immediate to realize that the correlation matrix BB^T depends on the parameters $\theta_{1,1}$ and $\theta_{2,1}$ only through their difference. One of the two angles, therefore, turns out to be redundant.

More generally, the degeneracy can be understood when we think of the geometric interpretation of the parameterization: we are seeking n unit vectors whose projections onto one another mimic the correlations we want. However, suppose we found any such set of vectors $\{\mathbf{v}_i\}$. It is intuitive that, if we rigidly rotate this bunch of vectors in *any* way in the n -dimensional space, we get another set of vectors *exactly equivalent to the original* in terms of mutual projections. This explains the origin of the degeneracy in the parameterization of Eq. (2), and suggests a modification that, breaking the rotational symmetry inherent in the problem, yields a more parsimonious parameterization still preserving its full flexibility.

Breaking up the symmetry: the triangular angles parameterization

If $\mathbf{x}^{(n)}$ is any point lying on the surface of a unit n -dimensional sphere, denote by $\mathbf{x}^{(n,k)}$ its projection onto the k -dimensional sphere in the (k -dimensional) subspace generated by $\mathbf{e}_{n-k+1}, \dots, \mathbf{e}_n$. The following recursive relation holds, where $\mathbf{x}^{(n,n)} = \mathbf{x}^{(n)}$ and $\mathbf{x}^{(n,1)} = (1)$,

$$(8) \quad \mathbf{x}^{(n,k)} = \cos \theta_{n,n-k+1} \begin{pmatrix} 1 \\ \mathbf{0}_{k-1} \end{pmatrix} + \sin \theta_{n,n-k+1} \begin{pmatrix} 0 \\ \mathbf{x}^{(n,k-1)} \end{pmatrix} = \begin{pmatrix} \cos \theta_{n,n-k+1} \\ \sin \theta_{n,n-k+1} \mathbf{x}^{(n,k-1)} \end{pmatrix}$$

with $\theta_{n,n-1} \in [0, 2\pi]$ and $\theta_{n,n-k+1} \in [0, \pi]$ for each $k = 3, \dots, n$, and where in general $\mathbf{0}_m = (0, 0, \dots, 0)^T$ (m components).

In the SAP of correlations all that counts are the angles between couples of vectors \mathbf{v}_i . For this reason the most natural way to break the symmetry is to start by choosing vector \mathbf{v}_1 (the first column of matrix B^T) equal to \mathbf{e}_1 , thereby eliminating the first source of redundancy, *i.e.* angles $\theta_{1,1}, \theta_{1,2}, \dots, \theta_{1,n-1}$.

Vector \mathbf{v}_2 can then be chosen lying in the (x_1, x_2) plane, so as to form an angle $\theta_{2,1} = \arccos(\rho_{1,2})$ with the first vector. It is intuitive that such a parameterization of the second vector, although sufficient to reconstruct the desired correlation $\rho_{1,2} = \mathbf{v}_1 \cdot \mathbf{v}_2$, is still redundant: in fact, had we chosen $\theta'_{2,1} = 2\pi - \theta_{2,1}$ the scalar product between \mathbf{v}_1 and \mathbf{v}_2 would again be $\rho_{1,2}$. In order to eliminate this redundancy it is sufficient to limit ourselves to choosing \mathbf{v}_2 from the upper halfplane, *i.e.* $\theta_{2,1} \in [0, \pi]$ (see Fig. 1).

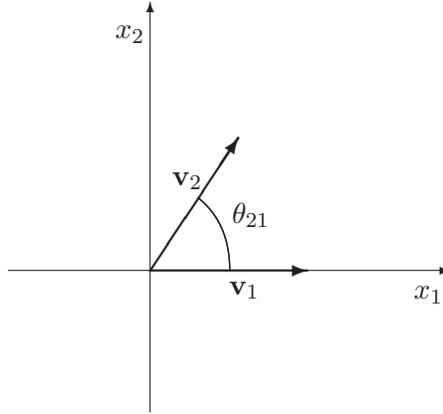


Figure 1: The first two vectors; note that, in order to break the degeneracy, $\theta_{2,1}$ must lie between 0 and π .

Vector \mathbf{v}_3 must satisfy two equalities concerning its projections onto \mathbf{v}_1 and \mathbf{v}_2 :

$$(9) \quad \begin{cases} \mathbf{v}_3 \cdot \mathbf{v}_1 = \rho_{1,3} \\ \mathbf{v}_3 \cdot \mathbf{v}_2 = \rho_{2,3} \end{cases}$$

which can be fulfilled by rotating vector \mathbf{e}_1 counterclockwise through two angles $\theta_{3,1}$ and $\theta_{3,2}$ in planes (x_1, x_2) and (x_2, x_3) , respectively. In order to span all possible values of two-body

correlations, still eliminating degeneracies, it is necessary to let the tip of vector \mathbf{v}_3 vary only over the upper hemisphere, thus having $\theta_{3,1}, \theta_{3,2} \in [0, \pi]$ (see Fig. 2).

This scheme can be repeated up to the last vector \mathbf{v}_n , the basic procedure being to add the generic k -th new vector \mathbf{v}_k through the specification of $k - 1$ angles. This vector will lie in the k -dimensional “upper” hemisphere of unit radius, which is accomplished by choosing all angles in $[0, \pi]$, as we will explain below.

Summarizing, the i -th row of matrix B is obtained by applying $i - 1$ consecutive counterclockwise Jacobi rotations to vector \mathbf{e}_1 of $\theta_{i,j}$ radians in (x_j, x_{j+1}) planes, $j = 1, \dots, i - 1$. With this procedure the actual number of free parameters (angles $\{\theta_{k,j}\}, j = 1, \dots, k - 1$ for vector \mathbf{v}_k) needed to describe all possible correlations is equal to $n(n - 1)/2$, as it should be.

Our matrix B , therefore, is defined by $b_{1,j} = \delta_{j1}$ and

$$(10) \quad b_{i,j} = \begin{cases} \cos \theta_{i,1} & \text{for } j = 1 \\ \cos \theta_{i,j} \prod_{k=1}^{j-1} \sin \theta_{i,k} & \text{for } 2 \leq j \leq i - 1 \\ \prod_{k=1}^{i-1} \sin \theta_{i,k} & \text{for } j = i \\ 0 & \text{for } i + 1 \leq j \leq n \end{cases}$$

for $i = 2, \dots, n$, and looks like the following:

$$(11) \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ \cos \theta_{2,1} & \sin \theta_{2,1} & 0 & 0 & \dots & 0 \\ \cos \theta_{3,1} & \cos \theta_{3,2} \sin \theta_{3,1} & \sin \theta_{3,2} \sin \theta_{3,1} & 0 & \dots & 0 \\ \cos \theta_{4,1} & \cos \theta_{4,2} \sin \theta_{4,1} & \cos \theta_{4,3} \sin \theta_{4,2} \sin \theta_{4,1} & \sin \theta_{4,3} \sin \theta_{4,2} \sin \theta_{4,1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Comparing (10) with (2), it is immediate to realize that we can retrieve our parameterization from the SAP by simply setting $\theta_{i,j} = 0$ for each $j \geq i$. Indeed, the θ 's in such a case form a triangular matrix, so that for example for $n = 4$ we obtain that the θ and B matrices are

$$\theta = \begin{bmatrix} 0 & 0 & 0 \\ \theta_{2,1} & 0 & 0 \\ \theta_{3,1} & \theta_{3,2} & 0 \\ \theta_{4,1} & \theta_{4,2} & \theta_{4,3} \end{bmatrix}, \quad B(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ b_{2,1} & b_{2,2} & 0 & 0 \\ b_{3,1} & b_{3,2} & b_{3,3} & 0 \\ b_{4,1} & b_{4,2} & b_{4,3} & b_{4,4} \end{bmatrix}.$$

Since these are both triangular matrices, we refer to this choice as the TAP.

The reason why angles $\{\theta_{i,j}\}, j = 1, \dots, i - 1$, for the generic vector \mathbf{v}_i , are constrained to vary in $[0, \pi]$ is explained by the existence of a bijective correspondence between unit-norm vectors in \mathbb{R}^n , with positive i -th component and last $n - i$ components equal to 0, and angles $\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,i-1} \in [0, \pi]^{i-1}$. In fact, denote by p_1, p_2, \dots, p_i the first i coordinates of a given point in \mathbb{R}^n lying on the i -dimensional “upper” hemisphere in the subspace generated by $\mathbf{e}_1, \dots, \mathbf{e}_i$ (positive i -th component). Then, there exists a unique choice of the angles in $[0, \pi]$ such that $b_{i,j} = p_j$, for each $j = 1, \dots, i - 1$, with $b_{i,j}$ defined as in (10). This statement is easily proved as follows.

For $j = 1$, we must have $\cos \theta_{i,1} = p_1$. There exists a unique angle in $[0, \pi]$ satisfying this equation, say $\bar{\theta}_{i,1}$. For $2 \leq j \leq i - 1$, we must have $\cos \theta_{i,j} \prod_{k=1}^{j-1} \sin \theta_{i,k} = p_j$. By induction on j , we can prove that $\prod_{k=1}^{j-1} \sin \bar{\theta}_{i,k}^2 = 1 - \sum_{k=1}^{j-1} p_k^2$, where $\bar{\theta}_{i,k}$ denotes the unique angle in $[0, \pi]$ that retrieves coordinate p_k , $2 \leq k \leq j - 1$. Since, by definition, $p_j^2 \leq 1 - \sum_{k=1}^{j-1} p_k^2$, the equation at step j admits a solution, which is unique provided that we choose $\theta_{i,j} \in [0, \pi]$. Finally, the i -th coordinate is matched by construction (unique positive number leading to a unit norm).

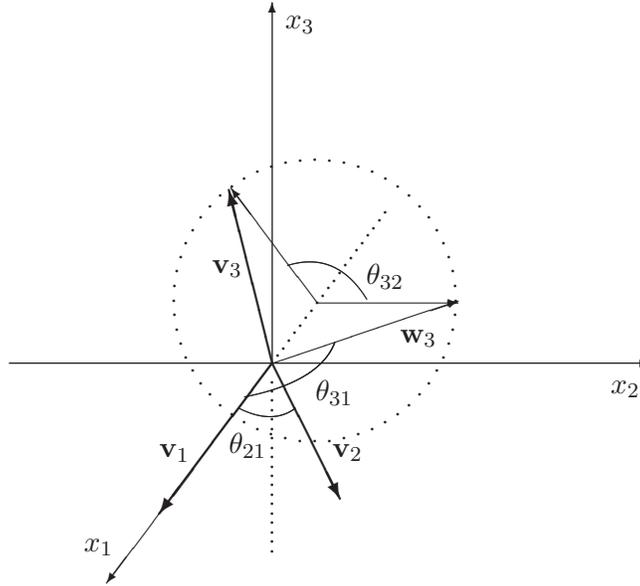


Figure 2: The first three vectors; in order to break the symmetry, still preserving the flexibility in the parameterization, $\theta_{2,1}, \theta_{3,1}, \theta_{3,2} \in [0, \pi]$. The (dotted) circle is the intersection between the plane of equation $x_1 = \cos \theta_{3,1}$ and the unit three-dimensional sphere. Vector \mathbf{v}_3 is obtained as follows: first rotate (counterclockwise) \mathbf{v}_1 by an angle $\theta_{3,1}$ on the (x_1, x_2) plane (vector \mathbf{w}_3), then rotate (counterclockwise) the projection of \mathbf{w}_3 onto the plane $x_1 = \cos \theta_{3,1}$ by an angle $\theta_{3,2}$.

The final parameterization: parameterizing the angles themselves

Symmetry breaking is useful for two reasons: it eliminates redundancies and defines tighter bounds for the variability of the optimization parameters (the angles). Accordingly, our parameterization speeds up considerably the minimization of a cost function such as (7) and usually prevents it from getting stuck in sub-optimal regions.

The optimization algorithms that are implemented in practice, however, are typically uncon-

strained.² In order to exploit the information on angle boundaries without explicitly constraining the angles to vary within their natural limits (*i.e.* $\theta_{k,j} \in [0, \pi]$ for $j = 1, \dots, k-1$), it is sufficient to re-parameterize them in terms of real numbers in $(-\infty, \infty)$ as follows:

$$(12) \quad \theta_{k,j} = \frac{\pi}{2} - \arctan x_{k,j} \quad \text{for } j = 1, \dots, k-1$$

The resulting matrix B can then be simplified through basic trigonometric equalities.

The angle parameterization *vs* a Cholesky decomposition

The Cholesky decomposition is a well known result in linear algebra stating the possibility of calculating the “square root” of a symmetric positive-definite matrix, which is uniquely factored in terms of a lower triangular matrix and its transpose. Formally [3] the following theorem holds.

Theorem 1. *For any symmetric positive-definite matrix $A \in \mathbb{R}^{n \times n}$ there exists a unique lower triangular matrix $L \in \mathbb{R}^{n \times n}$ with positive diagonal entries such that $A = LL^T$.*

Applying this result to our correlation matrix C (in the full-rank case), we can state the existence of a unique lower triangular matrix $L = (l_{i,j})_{i,j=1,\dots,n}$, with positive diagonal entries, such that, for any $j \geq i$,

$$(13) \quad \rho_{i,j} = \sum_{k=1}^i l_{i,k} l_{j,k}.$$

Remembering the previous results, we immediately have the following.

Theorem 2. *If matrix C is positive definite (full rank), our matrix B of Eq. (10) coincides with the Cholesky factorization L of C .*

In fact, the ultimate result of our procedure above is a lower triangular matrix with positive diagonal (the last component of every vector is by construction positive, in order to break the degeneracy of the parameterization of Eq. (2)). The Cholesky procedure can itself be interpreted in the light of the geometric argument exposed above: specializing to the case of correlation matrices, a close analysis reveals that [3, 6]

$$(14) \quad \left\{ \begin{array}{ll} \rho_{1,1} = l_{1,1}^2 & \Rightarrow l_{1,1} = \frac{1}{\sqrt{\rho_{1,1}}} = 1 \\ \rho_{1,2} = l_{1,1} l_{2,1} & \Rightarrow l_{2,1} = \rho_{1,2} \sqrt{\rho_{1,1}} = \rho_{1,2} \\ \rho_{1,3} = l_{1,1} l_{3,1} & \Rightarrow l_{3,1} = \rho_{1,3} \sqrt{\rho_{1,1}} = \rho_{1,3} \\ \rho_{2,2} = l_{2,1}^2 + l_{2,2}^2 & \Rightarrow l_{2,2} = \sqrt{\rho_{2,2} - \rho_{1,2}^2} = \sqrt{1 - \rho_{1,2}^2} \\ \rho_{2,3} = l_{2,1} l_{3,1} + l_{2,2} l_{3,2} & \Rightarrow l_{3,2} = \frac{\rho_{2,3} - l_{2,1} l_{3,1}}{l_{2,2}} = \frac{\rho_{2,3} - \rho_{1,2} \rho_{1,3}}{\sqrt{1 - \rho_{1,2}^2}} \\ \vdots & \vdots \end{array} \right.$$

and so on and so forth. This establishes a useful hierarchy in the entries of the Cholesky matrix, allowing for their direct recursive computation.

Denoting by \mathbf{u}_i the i -th row of matrix L , the structure of the matrix is obvious:

²A typical example is the widely used Levenberg-Marquardt method applicable to cost functions of type (7).

- \mathbf{u}_1 is nothing but vector \mathbf{e}_1^T ;
- \mathbf{u}_2 is the only unit vector in the (x_1, x_2) upper half plane such that $\mathbf{u}_2 \cdot \mathbf{u}_1 = \rho_{1,2}$: for this reason, $(\mathbf{u}_2)_1$ equals $\rho_{1,2}$ and $(\mathbf{u}_2)_2$ (the only nonzero component left for this vector) is chosen to be positive and so that the resulting vector has unit norm.
- \mathbf{u}_3 is the only unit vector in the (x_1, x_2, x_3) upper half space such that $\mathbf{u}_3 \cdot \mathbf{u}_1 = \rho_{1,3}$ and $\mathbf{u}_3 \cdot \mathbf{u}_2 = \rho_{2,3}$. Since $\mathbf{u}_3 \cdot \mathbf{u}_1 = \rho_{1,3}$, $(\mathbf{u}_3)_1$ equals $\rho_{1,3}$. Moreover, $\mathbf{u}_3 \cdot \mathbf{u}_2 = \rho_{2,3}$ implies that $(\mathbf{u}_3)_2 = (\rho_{2,3} - \rho_{1,2}\rho_{1,3})/\sqrt{1 - \rho_{1,2}^2}$. Finally, $(\mathbf{u}_3)_3$ is positive and such that the vector has unit length.

This scheme continues, if possible, row by row,³ thus building the triangular matrix L . The result (for a positive-definite matrix C) is the same we obtain when using the reduced form scheme. This happens because the philosophy is the same. In fact, every time, one: i) adds a vector to a preexisting set of unit vectors; ii) makes sure that projections of the last vector onto the preceding ones reproduce the desired correlations; iii) chooses the last component so as to bring back a unit norm (if possible) *with positive sign* (ensuring that this vector lies in the upper hemisphere).

The equivalence between our TAP method and the Cholesky procedure in case of positive definiteness, already remarked in [5], suggests a new optimization method to be applied when the matrix \hat{C} is not positive definite. The method corresponds to choosing the components of \mathbf{v}_k so as to generate projections between \mathbf{v}_k and all other vectors *as close as possible* to the desired correlations,⁴ and is based on the following modified Cholesky procedure. **should we keep this? Numbers are wrong!**

We proceed with the traditional Cholesky algorithm moving along rows as usual, up to when it works (say up to row \bar{k}). Every time a new row is completed, a check must be made that the corresponding vector can have unit length (in other words, that its last component is not imaginary). If this does not happen, a simple optimization based on angles $\theta_{\bar{k},j}$ can be done, so that the projections of vector $\mathbf{v}_{\bar{k}}$ onto the preceding ones are reproduced as close as possible.

Of course, this procedure cannot generate any better minimum than the global one previously outlined.

Reduced rank parameterizations

A reduced rank version of the decomposition is the following: suppose $C \in \mathbb{R}^{n \times n}$ is a correlation matrix that we want to decompose in an approximate, lower-rank way as

$$(15) \quad C \simeq BB^T$$

with $B \in \mathbb{R}^{n \times r}$, $1 < r \leq n$. This is a rank- r approximate decomposition of matrix C [2, 7].

One could still choose $B \in \mathbb{R}^{n \times n}$, setting the last $n - r$ columns of B equal to zero. It is evident that a rank- r parameterization within this scheme consists in choosing vectors $\{\mathbf{v}_k\}$ to lie within

³In common implementations of the algorithm, see for instance [6], a check is generally made that the square of the last nonzero entry of a row vector is actually positive.

⁴Here, as close as possible should read “as close as possible according to the particular choice of the cost function” of which Eq. (7) is but an example.

the r -dimensional (x_1, \dots, x_r) subspace of \mathbb{R}^n . We thus define $b_{1,j} = \delta_{j1}$ and

$$(16) \quad b_{i,j} = \begin{cases} \cos \theta_{i,1} & \text{for } j = 1 \\ \cos \theta_{i,j} \prod_{k=1}^{j-1} \sin \theta_{i,k} & \text{for } 2 \leq j < \min(i, r) \\ \prod_{k=1}^{j-1} \sin \theta_{i,k} & \text{for } j = \min(i, r) \\ 0 & \text{for } \min(i, r) < j \leq n \end{cases}$$

for $i = 2, \dots, n$. This is obtained from (10) by setting $\theta_{i,r} = 0$ for each $i > r$, with the subsequent angles $\theta_{i,r+1}, \dots, \theta_{i,i-1}$ that can be freely chosen.

A practical example helps in finding how many parameters are involved: the structure of a rank-3 matrix B is

$$(17) \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \cos \theta_{2,1} & \sin \theta_{2,1} & 0 & 0 & \cdots & 0 \\ \cos \theta_{3,1} & \cos \theta_{3,2} \sin \theta_{3,1} & \sin \theta_{3,2} \sin \theta_{3,1} & 0 & \cdots & 0 \\ \cos \theta_{4,1} & \cos \theta_{4,2} \sin \theta_{4,1} & \sin \theta_{4,2} \sin \theta_{4,1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

i.e. a triangle for the first three rows, on top of a $(n-3) \times 3$ rectangle of numbers. In the general case of a rank- r parameterization of an $n \times n$ matrix, the appearance of matrix B will be that of a triangle over the first r rows (corresponding to $r(r-1)/2$ free parameters) on top of a $(n-r) \times r$ rectangle (corresponding to $(r-1)(n-r)$ parameters). The total number of parameters is therefore $(r-1)(n - \frac{r}{2})$. For example, again in the case $n = 4$ but with rank $r = 3$ this time we have

$$\theta = \begin{bmatrix} 0 & 0 \\ \theta_{2,1} & 0 \\ \theta_{3,1} & \theta_{3,2} \\ \theta_{4,1} & \theta_{4,2} \end{bmatrix}, \quad B(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ b_{2,1} & b_{2,2} & 0 \\ b_{3,1} & b_{3,2} & b_{3,3} \\ b_{4,1} & b_{4,2} & b_{4,3} \end{bmatrix}.$$

Since both the θ and B matrices nonzero entries describe right trapezoids, we refer to this choice as to the “right trapezoid angles parameterization” (RTAP).

In the general case, $(r-1)(n - \frac{r}{2})$ is exactly the number of free parameters for a rank- r correlation matrix C . In the representation $\rho_{i,j} = \mathbf{v}_i \cdot \mathbf{v}_j$, the assumption that matrix C has rank $r \leq n$ implies that the last $n-r$ rows of C can be written as linear combinations of the first r rows (possibly upon recombination of rows and columns of C). One can thus write, for each $i > r$,

$$(18) \quad \rho_{i,j} = \sum_{k=1}^r \lambda_{i,k} \rho_{k,j}.$$

There are r parameters $\{\lambda_{i,k}\}$ for each line $i > r$. However, since $\rho_{i,i} = 1 = \sum_{k=1}^r \lambda_{i,k} \rho_{k,i}$ we have only $r-1$ free parameters for each of the remaining $n-r$ lines. The first r lines allocate a total of $r(r-1)/2$ free parameters (the entries $\rho_{i,j}$), thereby returning the total number of parameters one expects.

A full rank subparameterization for the Libor Market Model

For specific models a given behaviour for the correlation matrix should be imposed. The LMM is one such case, see also [1]. In fact, due to the extremely homogeneous nature of the simulated processes (they are all forward Libor rates spanning consecutive time intervals) some aspects of their instantaneous correlation structure are both intuitive and supported by historical evidence [7], [9]:

- B1) $\rho_{i,j} \geq \rho_{i,j+1}$ for any $j \geq i$ (correlation between rates should decrease when the interval between fixing dates increases)
- B2) $\rho_{i,j} \leq \rho_{i,j+1}$ for any $j < i$ (by the same token)
- B3) $\rho_{i,i+k}$ is an increasing function of i for each given $k \geq 1$ (e.g. correlation between Libor rates getting fixed in six months and two years from now is lower than correlation between Libors getting fixed in ten and a half years and twelve years from now, corresponding to the idea that the forward curve dynamics flattens for increasing maturities).

It is thus natural to wonder whether one can embed these properties in the parameterization of matrix C ; in fact much effort has been spent in this direction in recent literature [2, 9]. It turns out that this can indeed be accomplished within our geometrical interpretation, with few parameters, and with no effort and no explicit constraint on the parameter values. In fact, it suffices to suitably subparameterize matrix B : choose vectors $\{\mathbf{v}_i\}$, $i = 2, \dots, n$, iteratively as

$$(19) \quad (\mathbf{v}_i)_j = \begin{cases} \sin \phi_{i-1} (\mathbf{v}_{i-1})_j & \text{for } j < i \\ \cos \phi_{i-1} & \text{for } j = i, \end{cases}$$

where $\phi_1, \dots, \phi_{n-1} \in (0, \pi/2]$. In this case, the actual number of parameters, the ϕ 's, is $n - 1$.

The geometric idea underlying this new parameterization can be described as follows: take vector $\mathbf{v}_1 = \mathbf{e}_1$, \mathbf{v}_2 is obtained from \mathbf{v}_1 with a counterclockwise rotation by $\frac{\pi}{2} - \phi_1$ in the (x_1, x_2) plane; \mathbf{v}_3 is obtained by ‘‘tilting’’ \mathbf{v}_2 along the x_3 direction, and so on and so forth.

For instance, a 4×4 version of matrix B would look like

$$(20) \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \sin \phi_1 & \cos \phi_1 & 0 & 0 \\ \sin \phi_2 \sin \phi_1 & \sin \phi_2 \cos \phi_1 & \cos \phi_2 & 0 \\ \sin \phi_3 \sin \phi_2 \sin \phi_1 & \sin \phi_3 \sin \phi_2 \cos \phi_1 & \sin \phi_3 \cos \phi_2 & \cos \phi_3 \end{pmatrix}$$

to be compared with our general parameterization

$$(21) \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \cos \theta_{2,1} & \sin \theta_{2,1} & 0 & 0 \\ \cos \theta_{3,1} & \cos \theta_{3,2} \sin \theta_{3,1} & \sin \theta_{3,2} \sin \theta_{3,1} & 0 \\ \cos \theta_{4,1} & \cos \theta_{4,2} \sin \theta_{4,1} & \cos \theta_{4,3} \sin \theta_{4,2} \sin \theta_{4,1} & \sin \theta_{4,3} \sin \theta_{4,2} \sin \theta_{4,1} \end{pmatrix}$$

The correlation matrix implied by (20) is immediately obtained by noting that

$$(22) \quad \rho_{i,j} = \mathbf{v}_i \cdot \mathbf{v}_j = \prod_{k=i}^{j-1} \sin \phi_k.$$

There is a direct correspondence between the new parameterization (Eq. (20)) and a constrained parametrization of the earlier kind (Eq. (21)). In fact, for any $i > 1$,

$$(23) \quad \mathbf{v}_i \cdot \mathbf{v}_1 = \prod_{j<i} \sin \phi_j = \cos \theta_{i,1} \iff \theta_{i,1} = \arccos \prod_{j<i} \sin \phi_j$$

where, in particular, $\theta_{2,1} = \frac{\pi}{2} - \phi_1$. Moreover, for any $i > 2$,

$$(24) \quad \begin{aligned} \mathbf{v}_i \cdot \mathbf{v}_2 &= \left(\prod_{j<i} \sin \phi_j \right) \sin \phi_1 + \cos \phi_1^2 \left(\prod_{j=2}^{i-1} \sin \phi_j \right) \\ &= \cos \theta_{i,1} \cos \theta_{2,1} + \cos \theta_{i,2} \sin \theta_{i,1} \sin \theta_{2,1} = \left(\prod_{j<i} \sin \phi_j \right) \sin \phi_1 + \cos \theta_{i,2} \sin \theta_{i,1} \sin \theta_{2,1} \iff \\ &\iff \cos \theta_{i,2} \sin \theta_{i,1} \sin \theta_{2,1} = \cos^2 \phi_1 \left(\prod_{j=2}^{i-1} \sin \phi_j \right). \end{aligned}$$

If either $\sin \theta_{i,1} = 0$ or $\sin \theta_{2,1} = 0$, in which cases also the RHS is zero, we may choose $\theta_{i,2} = 0$; otherwise, noting that $\sin \theta_{2,1} = \cos \phi_1$,

$$(25) \quad \theta_{i,2} = \arccos \left(\frac{\cos \phi_1 \prod_{j=2}^{i-1} \sin \phi_j}{\sin \theta_{i,1}} \right).$$

The story continues in a recursive way. Therefore, for any parameterization of the kind of Eq. (19) one can find suitable angles $\{\theta_{i,j}\}$ in terms of which matrix B can be rewritten in the more traditional way. This parameterization is based on $n - 1$ parameters.

Proposition 3. *The parameterization of Eq. (19) describes the decomposition of a correlation matrix satisfying conditions B1) to B3) if and only if $\phi_1, \dots, \phi_{n-1}$ is an increasing sequence, namely $0 < \phi_i \leq \phi_{i+1} \leq \pi/2$ for each $i = 1, \dots, n - 2$.*

Proof. The first two properties trivially follow from (22). To prove the third, we remember that $\phi_i \in (0, \pi/2]$ so that $\sin \phi_i > 0$ for each i . Therefore,

$$\rho_{i,i+k} = \prod_{j=i}^{i+k-1} \sin \phi_j,$$

is increasing in i if and only if its logarithm is increasing.

$$(26) \quad \begin{aligned} \ln \rho_{i,i+k} &= \sum_{j=i}^{i+k-1} \ln \sin \phi_j \\ \rho_{i,i+k} \leq \rho_{i+1,i+k+1} &\iff \ln \sin \phi_i \leq \ln \sin \phi_{i+k} \iff \sin \phi_i \leq \sin \phi_{i+k}. \end{aligned}$$

Now, $\sin \phi_i \leq \sin \phi_{i+k}$ for each i and k if and only if $\sin \phi_i \leq \sin \phi_{i+1}$ for each i . Since $\phi_i \in (0, \pi/2]$ for each i , this is equivalent to say that the sequence $\phi_1, \dots, \phi_{n-1}$ must be increasing. \square

Remark 4. *Alternative approaches to the parameterization of a reasonable correlation matrix for the LMM can still be recast within a geometric view. In fact, it turns out that what was described in Proposition 3 is exactly equivalent to a parameterization proposed by Schoenmakers and Coffey [9], who parameterize the generic correlation $\rho_{i,j}$ through a set of n positive real numbers $1 = c_1 < c_2 < \dots < c_n$ with $\frac{c_1}{c_2} \leq \frac{c_2}{c_3} \leq \dots \leq \frac{c_{n-1}}{c_n}$ by setting $\rho_{i,j} = c_i/c_j$ for $i < j$. This can be seen to be equivalent to our parameterization by choosing $c_i = 1/(\prod_{j<i} \sin \phi_j)$ for each $i > 1$. Schoenmakers–Coffey’s condition that $\frac{c_1}{c_2} \leq \frac{c_2}{c_3} \leq \dots$ is equivalent to saying that $\sin \phi_i \leq \sin \phi_{i+1}$ for each i .*

Conclusions

The possibility to suitably parameterize a general correlation matrix is of fundamental importance in many financial applications. The angles parameterization proposed by Pinheiro and Bates [5] has been here reviewed and interpreted in an original fashion, based on a geometrical argument. Such geometrical interpretation allows us to derive a parameterization which is more parsimonious than that of Rebonato and Jäckel [8]. This is quite useful when seeking the correlation matrix that best approximates a given symmetric matrix, typically not positive semidefinite, or when solving issues related to a rank reduction in the correlation structure of a given basket of assets.

The intuitive understanding gained through our geometrical interpretation of the correlation matrix may often help in “engineering” correlation structures that embed given desirable features of the processes underlying an options market. One natural example comes from the Libor Market Model, where the stark relationship between rates spanning adjacent periods induces obvious features in their correlations. These features can be geometrically embedded in suitable subparameterizations of our angular scheme.

Acknowledgements

The authors are grateful to Walter Prah for helpful suggestions and for signalling the Pinheiro and Bates (1996) reference. They are also grateful to Aleardo Adotti, head of Product and Business Development at Banca IMI, for his constant support and encouragement.

References

- [1] C. Alexander, *Common Correlation Structures for Calibrating the LIBOR Model*, ISMA Centre Working Paper.
- [2] D. Brigo, F. Mercurio, *Interest Rate Models*, Springer Verlag (2001), updated section on correlation available at <http://www.damianobrigo.it/correl.pdf>
- [3] G.H. Golub, C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press (1996)
- [4] P. Jorion, *Value at Risk*, McGraw–Hill (2001)
- [5] J.C. Pinheiro, D.M. Bates, *Unconstrained Parameterization for Variance-Covariance Matrices*, *Statistics and Computing* (1996), 289-296
- [6] W.H. Press, S.A. Teukolsky, B.P. Flannery, *Numerical Recipes in C*, Cambridge University Press (1992)

- [7] R. Rebonato, *Volatility and Correlation*, Wiley (1999)
- [8] R. Rebonato, P. Jäckel, *The most general methodology to create a valid correlation matrix for risk management and option pricing purposes*, QUARC preprint, available at <http://www.rebonato.com> (1999)
- [9] J. Schoenmakers, B. Coffey, *Stable implied calibration of a multifactor Libor model via a semi-parametric correlation structure*, Weierstraß Institut für Angewandte Analysis und Stochastik preprint (2000)