Improving the most general methodology to create a valid correlation matrix

K. Schöttle & R. Werner
HVB-Stiftungsinstitut für Finanzmathematik, TU München, Germany,
risklab germany GmbH, Germany

Abstract

Jaeckel and Rebonato [1] develop two different methods of creating valid correlation matrices: construction by hypersphere decomposition and by singular value (i.e. spectral) decomposition. Although both methods yield satisfactory results in practice, from a mathematical point of view they both share some theoretical drawbacks.

Using results from semidefinite programming (SDP) we give the most general problem formulation to compute valid correlation matrices. We present numerical results which prove that these SDPs are rather easily solvable with efficient solvers for SDP problems (e.g. PENNON, see [3]). In contrast to Higham [2] we do not find numerical difficulties in solving the stated SDP problems.

We close the article with two more very important features which have been neglected in literature so far. First, regularity and second, control of the condition number of the resulting correlation matrix are easily guaranteed by linear SDP-constraints. Numerical experiments show that these additional constraints do not harm the efficient numerical solution.
Keywords: correlation matrix, semidefinite programming, risk management.

1 Motivation

In their article [1] Jaeckel and Rebonato introduce the problem of *finding the best correlation matrix*. In this context, a correlation matrix is given by the following

Definition 1. *Let $\boldsymbol{S}^n$ be the space of real symmetric $n \times n$ matrices. Further, let $\boldsymbol{S}^n_+ \subset \boldsymbol{S}^n$ be the cone of positive semidefinite matrices. A matrix $G \in \boldsymbol{S}^n_+$ ($G \succeq 0$) is called* correlation matrix*, if $G_{ii} = 1$ for $i = 1, \ldots, n$.*



Risk Analysis IV, C. A. Brebbia (Editor)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1

It is well known among practitioners that there are several situations when an estimated correlation matrix is in fact not a true correlation matrix. For example, correlations may be estimated from different data sources, specific entries may be manually altered for stress tests, rounding errors occur due to numerical difficulties, or some functional form of a correlation matrix from a fitting process fails to yield a true correlation matrix. In these situations, the candidate matrix C can be assumed to be symmetric and to have diagonal entries equal to one. The only violation is founded in the existence of some negative eigenvalue. For a more detailed motivation of the occurence of these *bad correlation matrices* we refer the reader to the article of Jaeckel and Rebonato [1].

1.1 The model

From their presentation it becomes obvious (although not explicitly stated) that their objective – together with the objective of almost all practitioners – lies in the solution of the following generic optimization problem:

$$\begin{aligned} \min_{G \in \mathbf{S}^n} \quad & f_C(G) \\ \text{s.t.} \quad & G \succeq 0, \\ & G_{ii} = 1, \quad i = 1, \ldots, n. \end{aligned} \tag{GCP}$$

Here, the notation means the following:

- C – the estimated bad correlation matrix.
- $f_C : \mathbf{S}^n \to \mathbb{R}$ is some appropriate objective function, depending on the bad correlation matrix C.

Please note that for a convex objective function f_C problem (GCP) is a standard convex (semidefinite) problem. For a very good introduction to the theory of SDP we refer the reader to Vandenberghe and Boyd [4].

1.2 The objective function f_C

The choice of a suitable objective function strongly depends on the application in mind. For example, a risk manager is looking for a true correlation matrix which simply is as close to the original estimate as possible. In this case, the objective function is most naturally given by some distance measure, i.e. $f_C(G) = ||G - C||$ for some matrix norm on $\mathbf{S}^n$. The most appropriate norm is the standard trace norm (also known as Frobenius norm, eigenvalue norm or 2-norm) $||G||_{\text{tr}}^2 := \langle G, G \rangle = \text{tr}(GG)$, making the $\mathbf{S}^n$ a Hilbert space with the scalar product $\langle A, B \rangle = \text{tr}(AB)$. If not explicitly stated otherwise, we assume in the following that $f_C(G) = ||G - C||_{\text{tr}}$.

A different choice of the objective function f_C may be the calibration error for some correlation matrix G, where G is e.g. obtained by some interest rate model calibration process (for an excellent treatment of this specific problem see e.g. the working paper by Brace and Womersley [5] or d'Aspremont [6]). In this situation,

Risk Analysis IV, C. A. Brebbia (Editor)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1

the function f_C may be non-convex in G. We will show later how we can deal with such problems in an easy way using a general purpose SDP solver.

2 Criticism on existing approaches

In the following paragraphs we will briefly summarize already existing methods which approximately solve the above optimization problem (GCP) together with some short criticism on each of them.

We have illustrated the geometrical interpretation of the shrinkage approach (named $G(\lambda^*)$), the spectral decomposition (C^{++}) and the general SDP method (G^*) in Figure 1 for an initial bad correlation matrix C.

2.1 Shrinkage

The main idea of the shrinkage approach is as follows. Additionally to the estimated bad correlation matrix C one chooses a target correlation matrix C_0 (as reference, see e.g. Kupiec, [7]). Restricting the solution of (GCP) to the line between C and C_0, problem (GCP) becomes in this case

$$\begin{aligned} \min_{\lambda \in \mathbb{R}} \quad & \lambda \\ \text{s.t.} \quad & G(\lambda) := C + \lambda(C_0 - C) \succeq 0. \end{aligned}$$

Please note that

$$||G(\lambda) - C||_{\text{tr}} = ||\lambda(C_0 - C)||_{\text{tr}} = \lambda ||C_0 - C||_{\text{tr}}$$

and hence the above objective function is equivalent to the original one.

Of course, it does not make sense to use a general purpose SDP solver for this easy problem. Kupiec uses some simple one-dimensional root finding algorithm together with a complete matrix orthogonalization at each iteration to control positive semidefiniteness. It would be numerically more robust and more efficient to test on positive semidefiniteness by Cholesky decomposition or to compute the smallest eigenvalue of the current iterate.

However, although the method is very easy to implement and works well even for large matrices, the dependence on the (artificial) target matrix C_0 is a major drawback in practice and disqualifies it for practical use.

2.2 The hypersphere decomposition

The hypersphere decomposition approach relies on the observation that for a suitably chosen matrix B (motivated by spherical coordinates) an approximation of C can be given by BB^T. The definition of an appropriate B is as follows:

$$B_{ij}(\Theta) := \begin{cases} \cos(\Theta_{ij}) \prod_{k=1}^{j-1} \sin(\Theta_{ik}) & \text{for } j < n, \\ \prod_{k=1}^{j-1} \sin(\Theta_{ik}) & \text{for } j = n. \end{cases}$$



Risk Analysis IV, C. A. Brebbia (Editor)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1

One can easily show that BB^T is positive semidefinite and has diagonal entries equal to one. The main advantage of this approach is that now with $G = G(\Theta) = B(\Theta)B(\Theta)^T$ problem (GCP) can be written as

$$\min_{\Theta \in \mathbb{R}^{n \times n-1}} ||B(\Theta)B(\Theta)^T - C||_{\text{tr}}^2$$

which becomes an unconstrained optimization problem and can as such be efficiently solved using e.g. BFGS or Newton's method, as gradient and Hessian are easily (even analytically) computable. In contrast to the shrinkage approach, no matrix decompositions (which have complexity $\mathbf{O}(n^3)$) have to be computed. However, the assembly of the Hessian has complexity $\mathbf{O}(N^2)$ and the solution of the according linear system even is of order $\mathbf{O}(N^3)$ where $N = n^2 - n$ is the number of optimization variables.

Still, the main drawback of this optimization problem is not its high complexity. As the optimization variable Θ enters the objective in a non-convex fashion, classical optimization methods fail to compute the global optimum. Even for realistic small dimensions (i.e. $n = 10$) we already have to solve a global optimization problem with $n^2 - n$ (i.e. $N = 90$) variables. Although the computation of local minimizers (which is often sufficient in practice) can be done very efficiently, global optimization problems should be avoided if possible.

2.3 The spectral decomposition

The third approximation to (GCP), originally introduced by Jaeckel and Rebonato, lies in the following two simple observations

Definition 2. *Let $D \in \boldsymbol{S}^n$ be a diagonal matrix. Then the diagonal matrix D^+ defined by $(D^+)_{ii} := \max(D_{ii}, 0)$ $(i = 1, \ldots, n)$, is the* projection *of D onto the cone $\boldsymbol{S}_+^n$, i.e. it solves*

$$\min_{G \in \boldsymbol{S}_+^n} ||G - D||_{\text{tr}}^2.$$

This characterization can easily be extended to general matrices from $\mathbf{S}^n$, where we use the well-known fact that each real symmetric matrix $A \in \mathbf{S}^n$ has a *spectral decomposition* $A = U\Lambda U^T$ with an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ (eigenvectors) and a diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ (eigenvalues).

Lemma 3. *Let $A \in \boldsymbol{S}^n$, with $A \notin \boldsymbol{S}_+^n$. Then the following optimization problem*

$$\min_{G \in \boldsymbol{S}_+^n} ||G - A||_{\text{tr}}^2$$

is solved by $G^* = A^+$, where A^+ is defined through its spectral decomposition as $A^+ := U\Lambda^+ U^T$. In this sense, A^+ can be seen as the projection of A onto $\boldsymbol{S}_+^n$. It holds that $(A^+)_{ii} \geq A_{ii}$ for $i = 1, \ldots, n$.

Proof. Using the invariance of the trace with respect to orthogonal transformation, the claim is obvious. $(A^+)_{ii} \geq A_{ii}$ follows from the fact that $A^+ \succeq A$. □

Risk Analysis IV, C. A. Brebbia (Editor)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1

Having a close look at (GCP) we see that the only difference to the optimization problem from the previous lemma lies in the constraints $G_{ii} = 1$. To take this constraint into account, the projection needs to be slightly adjusted.

Definition 4. *Let $A \in \boldsymbol{S}^n$, $A_{ii} > 0$. Then a diagonal scaling matrix $F = F(A)$ is defined by $F_{ii} := \left(A_{ii}\right)^{-\frac{1}{2}}$.*

Now we are ready to state the complete algorithm to compute a feasible approximation C^{++} to a bad correlation matrix C in the following theorem.

Theorem 5. *Let $C \in \boldsymbol{S}^n$, with $C \notin \boldsymbol{S}_+^n$ and $C_{ii} = 1$. Then the matrix C^{++} defined as*

$$C^{++} := F(C^+) \cdot C^+ \cdot F(C^+)$$

is feasible in (GCP), i.e. $C^{++} \in \boldsymbol{S}_+^n$ and $C_{ii}^{++} = 1$.

Proof. Due to the previous lemmas, the matrix $F(C^+)$ is well defined as $(C^+)_{ii} \geq C_{ii} = 1 > 0$. The rest is obvious. □

Please note that in general this matrix C^{++} is not necessarily an optimal solution to (GCP), thus we can only expect sub-optimality. However, we can expect C^{++} to be a very good approximation to the optimal solution of (GCP).

In contrast to the first two approaches, this method does not need the solution of any optimization problem. The main computation time is spent for the calculation of the spectral decomposition, i.e. the computation of the projection onto the cone of positive semidefinite matrices. The rest is simple scaling in the rows and columns of the resulting projection. From a complexity point of view this approach clearly outperforms all other methods due to its simplicity and efficiency. Only for very large matrices C we run into numerical problems computing a proper spectral decomposition.

2.4 Some common drawbacks

Although rather satisfying from a practical point of view, the previous approach shares two drawbacks with the other methods presented so far.

- From a mathematical point of view, all approaches only compute approximations to the original problem (GCP). The question how to solve (GCP) efficiently still remains open.
- It is rather hard to incorporate further constraints on the resulting improved correlation matrix. All methods rely on the specific problem formulation and cannot deal with additional constraints. We will deal with the problem of incorporating additional constraints in the last section.

Most important, due to optimality reasons, all methods end up with singular correlation matrices. As these matrices are e.g. further used in Markowitz optimization or the computation of value-at-risk quantities, regular matrices instead of singular ones are required. The shrinkage approach and the singular value decomposition can easily be changed to such a requirement, whereas it remains unclear how to extend the hypersphere decomposition.

Risk Analysis IV, C. A. Brebbia (Editor)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1

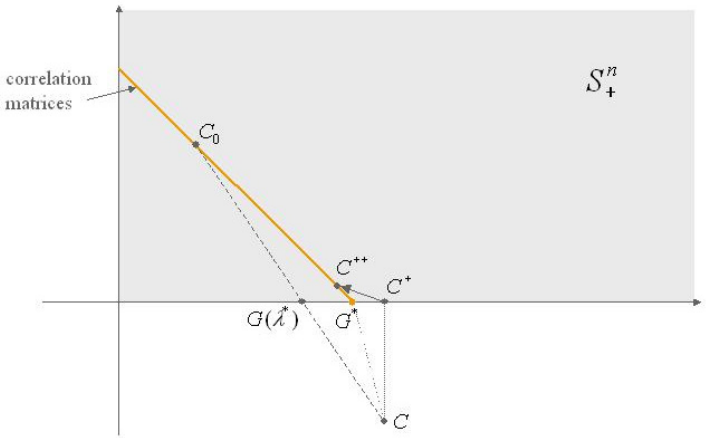


Figure 1: Geometrical interpretation of the estimated matrix C together with approximations.

3 Solution by efficient SDP solvers

In contrast to all methods proposed so far, we suggest the usage of standard SDP solvers. Current state-of-the-art SDP solvers are capable of solving medium sized problems (i.e. $50 \leq n \leq 100$) with acceptable computational effort, see Mittelmann's web page [8]. Using these standard solvers yields two strong advantages: First, we can easily include additional constraints (see next section). Second, the user only has to formulate (GCP) in the appropriate input format (e.g. standard SDPA format, see [9]) and can then use a variety of different solvers.

As problem (GCP) is not in standard SDP form, we have to reformulate (GCP) to fit the required input format.

3.1 Reformulations

Using calculation techniques from SDP theory (the Schur complement, see [10]), we can show

© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1

Theorem 6. *Problem (GCP) is equivalent to*

$$\begin{aligned} \min_{G,H \in \boldsymbol{S}_+^n} & \operatorname{tr}(H) \\ \textit{s.t.} \quad & \begin{pmatrix} H & G \\ G & I \end{pmatrix} \succeq \begin{pmatrix} 0 & C \\ C & 0 \end{pmatrix}, \\ & G_{ii} = 1, \quad i = 1, \ldots, n, \end{aligned}$$

or

$$\begin{aligned} \min_{G \in \boldsymbol{S}_+^n,\, h_{ij} \in \mathbb{R}} & \sum_{i,j} h_{ij} \\ \textit{s.t.} \quad & \begin{pmatrix} h_{ij} & G_{ij} - C_{ij} \\ G_{ij} - C_{ij} & 1 \end{pmatrix} \succeq 0, \quad 1 \le i \le j \le n, \\ & G_{ii} = 1, \quad i = 1, \ldots, n. \end{aligned}$$

Using an appropriate choice of the basis in the matrix space, the constraints on the diagonal elements of G are already implicitly fulfilled.

Remark 7. *A similar theorem holds for* $f_C(G) = ||G - C||_1 = \sum_{i,j} |G_{ij} - C_{ij}|$ *which is numerically more suited, especially for larger* n.

3.2 Example

The following illustrative example has been computed with PENNON, using a version of (GCP) with $f_C(G) = ||G - C||_1$. Please compare this singular matrix with the optimal matrix in the following section, where we have included lower bounds on the eigenvalues. For

$$C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \text{ we have } G^* = \begin{pmatrix} 1 & 0.7071 & 0 \\ 0.7071 & 1 & 0.7071 \\ 0 & 0.7071 & 1 \end{pmatrix}.$$

The original matrix C has the eigenvalues $-0.4142, 1, 2.4142$ and the optimal G^* has the eigenvalues $0, 1, 2$.

3.3 Generalization to non-convex problems

So far we have only considered problem (GCP) with $f_C(G) = ||G - C||$. If we are interested in the solution of calibration problems as arise e.g. in interest rate models, the objective is no longer convex. In this case none of the approximation methods is suited to solve these problems. Using SDP solvers tailored for non-convex problems (e.g. PENNON) we can hope to get at least local solutions. Preliminary tests are promising.



Risk Analysis IV, C. A. Brebbia (Editor)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1

3.4 Remarks on Higham's SDP approach

Higham claims in [2] that the problem of finding the best correlation matrix is not solvable with the help of semidefinite programming. In his context, he reformulates the original problem (using the Schur complement) as a SDP with $\mathbf{O}(n^4)$ constraints. He is completely right when claiming that problems like this can hardly be solved. However, when using the equivalent formulation from the previous paragraphs it is possible to formulate SDPs in standard form with $N = n^2$ variables with either only one $2n \times 2n$ and two $n \times n$ matrix constraints or with one $n \times n$ matrix constraint together with $\frac{n(n+1)}{2}$ 2×2 matrix constraints, which keeps the problem sizes within reasonable limits.

We have to admit that our method still is in the same complexity class as the hypersphere decomposition, i.e. we have complexity $\mathbf{O}(N^2)$ for the assembly of the Hessian and $\mathbf{O}(N^3)$ for the solution of the according linear system. Nevertheless, let us mention three points. First, when working with large matrices it may be more appropriate to use only first order information for the solution of (GCP). Second, in contrast to Higham, we have never come across fitting problems with correlation matrices with $n \geq 100$ where C did not possess a very special (sparse, block or functional dependence) structure in practice. Third, when working with large matrices, condition numbers are getting more and more important for subsequent computations and, in our opinion, should be taken into account. We completely agree with Higham that the *best* methodology to solve (GCP) in its most general form even for large matrices remains an open problem.

4 Including further constraints

As we have already mentioned, applications in stress testing or asset management often lead to the problem of incorporating additional (linear equality or inequality) constraints on certain elements of the correlation matrix, i.e. (GCP) is e.g. refined to

$$\begin{aligned} \min_{G \in \mathrm{S}^n} \quad & f_C(G) && \text{(GCP')} \\ \text{s.t.} \quad & G \succeq 0, \\ & G_{ii} = 1, \quad i = 1, \ldots, n, \\ & G_{jk} = G_{kj} = p_{jk}, \quad (j,k) \in \mathbf{J} \end{aligned}$$

for certain entries $(j,k) \in \mathbf{J}$ and values p_{jk}. In the setting of semidefinite programming, these constraints can rather easily be incorporated. In the same way we can introduce linear inequalities on the matrix elements, without harming the linear structure of the semidefinite programming problem.

As we have already mentioned, we do not want to obtain a singular matrix G^*, but a regular one. This can be guaranteed by the further (linear) constraint $G \succeq \varepsilon I$. As we can see from Table 1, this regularity constraint does not significantly change the computation time, in contrast, computation time sometimes decreases due to a



Risk Analysis IV, C. A. Brebbia (Editor)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1

Table 1: Computation times for (GCP) in seconds with and without regularity constraints on a standard notebook using PENNON.

problem size	20	30	50	75
without constraints	< 1	4	132	1322
with constraints	< 1	4	104	1379

better conditioning of the problem. If we recompute the optimal matrix from the previous example with a lower bound of 0.1 on the eigenvalues of G, we obtain

$$G^* = \begin{pmatrix} 1 & 0.6364 & 0 \\ 0.6364 & 1 & 0.6364 \\ 0 & 0.6364 & 1 \end{pmatrix}$$

which now has eigenvalues $0.1, 1.0, 1.9$. Please note that the optimal objective value increases from 0.5858 to 0.7272 when imposing the regularity constraint.

Another, even more important constraint can be imposed on the condition number of the improved correlation matrix. In normal circumstances, the condition number should be bounded above by α, i.e. the user wants to guarantee

$$\text{cond}(G) := \frac{\lambda_{\max}(G)}{\lambda_{\min}(G)} \leq \alpha.$$

Here, $\lambda_{\max}(G)$ and $\lambda_{\min}(G)$ denote the maximal and minimal eigenvalue of G. We can reformulate this constraint by

$$\lambda_{\max}(G) \leq \alpha \cdot \lambda_{\min}(G).$$

From the theory of semidefinite programming we know that we can add this constraint to (GCP) as

$$\begin{aligned} \min_{G \in \mathbf{S}^n,\ l,u \in \mathbb{R}} \quad & f_C(G) \\ \text{s.t.} \quad & G \succeq 0, \\ & l \cdot I_n \preceq G \preceq u \cdot I_n, \\ & u \leq \alpha l, \\ & G_{ii} = 1, \quad i = 1, \ldots, n. \end{aligned}$$

5 Summary

We have briefly presented well-known approaches to improve bad correlation matrices to suitable ones. We have shown that all approaches share two drawbacks

Risk Analysis IV, C. A. Brebbia (Editor)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1

which can be overcome by a semidefinite programming formulation. In contrast to other authors we do not run into numerical difficulties working with small and medium sized matrices. Using the very general semidefinite programming approach, we can easily include further constraints on elements or even on the condition number of the improved matrix.

References

[1] Jaeckel, P. & Rebonato, R., The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. *Journal of Risk*, **2(2)**, Winter 1999/2000.
[2] Higham, N., Computing the nearest correlation matrix – a problem from finance. *IMA Journal of Numerical Analysis*, **22**, pp. 329–343, 2001.
[3] Kocvara, M. & Stingl, M., PENNON – A Code for Convex Nonlinear and Semidefinite Programming. *Optimization Methods and Software* **18(3)**, pp. 317–333, 2003.
[4] Vandenberghe, L. & Boyd, S., Semidefinite Programming. *SIAM Review*, **38(1)**, pp. 49-95, 1996.
[5] Brace, A. & Womersley, R.S., Exact fit to the swaption volatility matrix using semidefinite programming. *Working paper*, ICBI Global Derivatives Conference, Paris, 2000.
[6] d'Aspremont, A., Interest Rate Model Calibration and Risk-Management Using Semidefinite Programming. *PhD Thesis*, Ecole Polytechnique, 2002.
[7] Kupiec, P.H., Stress testing in a value at risk framework. *Journal of Derivatives*, **6(1)**, pp. 7–24, 1998.
[8] Benchmarks for Optimization Software. http://plato.la.asu.edu/bench.html
[9] SemiDefinite Programming Algorithm. http://sdpa.is.titech.ac.jp
[10] Ben-Tal, A. & Nemirovski. A., Robust convex optimization. *Mathematics of Operations Research*, **23**, 1998.



Risk Analysis IV, C. A. Brebbia (Editor)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-736-1