OPTIMAL COVARIANCES IN RISK MODEL AGGREGATION

Alec N. Kercheval
Department of Mathematics
Florida State University
Tallahassee, FL 32306-4510
kercheval@math.fsu.edu

ABSTRACT
Abstract: Portfolio risk forecasts are often made by estimating an asset or factor covariance matrix. Practitioners commonly want to adjust a global covariance matrix encompassing several sub-markets by individually correcting the sub-market diagonal blocks. Since this is likely to result in the loss of positive semi-definiteness of the overall matrix, the off-diagonal blocks must then be adjusted to restore that property. Since there are many ways to do this adjustment, this leads to an optimization problem of Procrustes type. We discuss two solutions: a closed form solution using an adapted norm, and a fast majorization approach.

KEY WORDS
covariance matrix, risk forecast, orthogonal procrustes, matrix optimization.

1 Introduction

Equity portfolio risk forecasts are typically derived from a forecast of the asset or factor covariance matrix of returns. The following risk model aggregation problem is well-known to pension funds and mutual fund firms:

How can the firm evaluate the total risk of the combined portfolios of many managers?

Assume that we have $K > 1$ managers, each responsible for a portfolio in one of K markets. We also assume that each of these managers is able to construct with confidence a factor covariance matrix $\tilde{A}_k$ of size $n_k \times n_k$, $(k = 1, \ldots, K)$, which successfully describes the risk of portfolios in market k.

To understand the firmwide total risk of the union of the K managers' portfolios, we need a large covariance matrix $\tilde{V}$ of size $N \times N$, where $N = n_1 + \cdots + n_k$. The matrix $\tilde{V}$ must agree on its diagonal blocks with the $\tilde{A}_k$'s, and have meaningful information on the off-diagonal blocks about correlations between factors in different markets. Unfortunately, while there may be enough historical data to estimate covariance matrices of size $n_k \times n_k$, there almost surely is not enough data to estimate directly an $N \times N$ covariance matrix.

Instead, the firm's risk manager may do the following:

1. Develop a first-draft $N \times N$ covariance matrix V using the available data history, in whatever way seems to best capture cross-market correlations.

2. Replace the diagonal blocks A_k of V with the previously estimated market covariance matrices $\tilde{A}_k$ to form a new matrix $\tilde{V}$.

3. Since this action is likely to spoil positive semi-definiteness by creating negative eigenvalues, the off-diagonal blocks of $\tilde{V}$ must then be adjusted in some minimal way to restore positive definiteness.

This becomes an optimization problem involving positive definite matrices, which is the focus of this paper.

2 Simplifying the problem

In the most basic version of the problem, we are given a positive definite matrix V, expressed in block form as

$$V = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}.$$

This is intended to be a first draft global covariance matrix. Here A is a small diagonal block corresponding to covariances of the factors in one of the individual markets. We'll say that A is $n \times n$, V is $N \times N$, and for simplicity assume $n < N/2$.

Independently, we are also given a better estimate $\tilde{A}$ for the factor covariance matrix of that market. We wish to substitute $\tilde{A}$ for A in V, without changing the other factors described by C.

To avoid creating negative eigenvalues, this means we need to allow freedom to adjust B to restore positive semidefiniteness. The problem then becomes that of finding $\tilde{B}$ such that

$$\tilde{V}(\tilde{B}) = \begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{B}^T & C \end{pmatrix}$$

is positive semidefinite and as close as possible, in some suitable sense, to V.

3 Prior approaches

3.1 Rebonato and Jäckel

If we wish we can phrase this problem in terms of correlation matrices instead of covariance matrices, simply by normalizing the variables. In this context, Rebonato and Jäckel [10] considered the problem of the nearest $n \times n$ correlation matrix to a given symmetric matrix with unit diagonal, and proposed a solution involving minimizing a norm over an $n(n-1)$ dimensional parameter space. It's not difficult, e.g. [2], to improve this dimension by a factor of 2, but one can't avoid the problem that the solution will necessarily have zero eigenvalues. Also the nonconvex optimization problem becomes rapidly difficult as n grows.

3.2 Positive semidefinite programming

An improvement on the technique of Rebonato and Jäckel makes use of the concept of positive semidefinite programming, e.g. [5], [8]. Here, one notes that the space $\mathcal{S}$ of positive semidefinite matrices is a convex cone in the set of all $n \times n$ matrices, and prescribing diagonal blocks represents a simple linear constraint.

Hence the problem becomes a convex optimization problem, which can be solved for quite large n.

While this approach is very powerful, there are two difficulties in this context.

- Of necessity, if $\tilde{V}(B)$ has negative eigenvalues, then the optimum $\tilde{V}(\tilde{B})$ minimizing the norm

$$||\tilde{V} - V||$$

  will have zero eigenvalues, which is inconvenient for risk management applications. (One could add a further constraint that eigenvalues be larger than a certain chosen lower bound, but this is ad hoc.)

- Worse, the solution to the problem posed this way will always represent a change of the underlying variables that mixes factors across markets, which is financially undesirable. We explain this point in the next section.

4 Changing covariance matrices means changing variables

Notation. Let M_N denote the vector space of $N \times N$ real matrices, $GL(N, \mathbb{R})$ the subset of invertible matrices, and denote by $COV(N)$ the subset of all possible $N \times N$ covariance matrices of some N-dimensional random vector. (In our application, the random vector will be the vector of factor returns.) Equivalently, $COV(N)$ is the space of $N \times N$ positive semidefinite (symmetric) matrices. The subset of positive definite matrices will be denoted $COV^+(N)$.

The following fact is elementary:

Proposition 4.1 *If $V \in COV^+(N)$, then*

1. *$\{LVL^T : L \in M_N\} = COV(N)$, and*
2. *$\{LVL^T : L \in GL(N, \mathbb{R})\} = COV^+(N)$.*

Moreover, if V is the covariance matrix of a random vector s, the matrix LVL^T is the covariance matrix induced by the linear change of variables

$$\tilde{s} = Ls.$$

Therefore we may think of changing the covariance matrix V to a new matrix $\tilde{V}$ as equivalent to making a linear change of variables of the underlying factors s. Since factors are determined via linear regression on the asset returns, their identities are somewhat approximate in the first place – financially we can tolerate a small change L close to the identity as a correction in light of the exogenous information in $\tilde{A}$.

However, we need to preserve the identities of the individual markets corresponding to the diagonal blocks of V, and furthermore we don't want to touch factors outside the A block when making the above correction.

This means our change of variables should be constrained to the following block diagonal form:

$$L = \begin{pmatrix} L_1 & 0 \\ 0 & I \end{pmatrix}. \tag{1}$$

It is now not difficult to show the following [1]:

Proposition 4.2 *With V, $\tilde{V}$, and L as above,*

$$LVL^T = \tilde{V}$$

if and only if

$$L_1 = \tilde{A}^{1/2} O A^{-1/2},$$

where O is orthogonal and the exponent 1/2 refers to the unique positive definite square root.

Equivalently, the block diagonal constraint on L implies that the rectangular block $\tilde{B}$ is constrained to be of the form

$$\tilde{B} = \tilde{A}^{1/2} O A^{-1/2} B$$

for some orthogonal matrix O.

Notice now that when, as we have assumed, V, A, and $\tilde{A}$ are positive definite, then any admissible revised covariance matrix

$$\tilde{V}(O) = LVL^T$$

$$= \begin{pmatrix} \tilde{A} & \tilde{A}^{1/2} O A^{-1/2} B \\ B^T A^{-1/2} O^T A^{1/2} & C \end{pmatrix} \tag{2}$$

is necessarily invertible, because

$$L = \begin{pmatrix} \tilde{A}^{1/2} O A^{-1/2} & 0 \\ 0 & I \end{pmatrix} \tag{3}$$

is invertible (and with condition number bounded uniformly in the choice of O).

This is a natural resolution to the difficulty of zero eigenvalues when $\tilde{B}$ is unconstrained, as in the positive semidefinite programming approach. Note also that, because we have characterized the solutions satisfying the market integrity constraint (1), this means that any solution having zero eigenvalues must necessary fail to satisfy (1), and therefore represents a change of variables undesirably mixing factors across markets.

5 The optimization problem

We now are searching for the admissible $\tilde{V}$ closest to V, where "admissible" means $\tilde{V} = LVL^T$ for some L satisfying (3). In terms of some suitable norm, we want to minimize

$$||\tilde{V} - V||.$$

If we take the norm to be the usual Frobenius norm $||X||_F = tr(XX^T)$, this is equivalent to minimizing the off-diagonal block norm:

$$||\tilde{B} - B||_F = ||\tilde{A}^{1/2} O A^{-1/2} B - B||_F \quad (4)$$

as O varies over the orthogonal group $\mathrm{O}(n)$. This is an unconstrained but nonconvex problem of dimension $n(n-1)/2$, see [1].

There is a related classical problem in numerical linear algebra, e.g. [3]:

> *Orthogonal Procrustes Problem*: Given $m \times n$ matrices A and D, find $O \in \mathrm{O}(n)$ minimizing $\|AO - D\|_F$.

This has a closed form solution in terms of singular value decompositions (described in Section 6), but unfortunately there is no known exact solution to the problem we are facing, which we call the

> *Double Orthogonal Procrustes Problem*: Given matrices A, B, and D of compatible sizes, find an orthogonal matrix O minimizing $\|AOB - D\|_F$.

6 Choosing an adapted norm

Initial experiments in Anderson, et. al. [1] show that the numerical cost of minimizing the objective (4) over the orthogonal group $\mathrm{O}(n)$, via the standard Levenberg-Marquardt optimization routine, starts to become high once n is larger than around 20. We need to find a faster solution without giving up our constraint (1) on admissible changes of variables L.

The idea in this section is to take advantage of the freedom we have to choose our norm. By proper choice of norm, we can provide an exact closed form solution of our problem for which the only computation required is calculation of a single singular value decomposition.

Higham [5] describes a common weighted variant of the Frobenius norm:

$$||X||_W = ||W^{1/2} X W^{1/2}||_F$$

for some positive definite weighting matrix W. Often, W is chosen to be diagonal, but it need not be.

Consider now the following specific choice of W:

$$W = \begin{pmatrix} \tilde{A}^{-1} & 0 \\ 0 & I \end{pmatrix}.$$

We denote by $||.||_*$ the norm $||.||_W$ for this choice of W.

Our problem now is to minimize

$$||\tilde{V}(O) - V||_* \quad (5)$$

as O ranges over $\mathrm{O}(n)$, and where $\tilde{V}(O)$ is given by (2). Substituting our choice of W, this is equivalent to minimizing the quantity

$$||\tilde{A}^{-1/2}(\tilde{B} - B)||_F = ||OA^{-1/2}B - \tilde{A}^{-1/2}B||_F.$$

This is now subject to exact solution via the usual orthogonal procrustes method, as we now describe.

For convenience let $X = A^{-1/2}B$ and $Y = \tilde{A}^{-1/2}B$. We are minimizing

$$tr((OX - Y)(OX - Y)^T)$$
$$= tr(XX^T) + tr(YY^T) - 2tr(OXY^T)$$

as O varies over $\mathrm{O}(n)$. Since the first two terms don't depend on O, we are equivalently maximizing $tr(OXY^T)$.

Let UDV^T be the singular value decomposition of $XY^T = A^{-1/2}BB^T\tilde{A}^{-1/2}$. We want to find O maximizing

$$tr(OXY^T) = tr(OUDV^T) = tr(V^TOUD).$$

Notice V^TOU is orthogonal, D is diagonal with nonnegative entries. Some thought will convince the reader that the maximum occurs when $V^TOU = I$, or $O = VU^T$. This now minimizes the objective (6). The solution $\tilde{V}$ may thus be computed easily for any dimension for which the singular value decomposition is available.

The choice of the norm $||.||_*$ roughly amounts to giving equal weight to the principal components of the covariance block $\tilde{A}$. This is fine when the precise weightings are not too important in the application. However, the user may need to keep the Frobenius norm, or substitute some other specific weighting W. In this case, we have to address the full Double Orthogonal Procrustes Problem, as in the next section.

7 Koschat-Swayne Majorization

The Double Orthogonal Procrustes Problem is difficult because it is a nonconvex, high-dimensional problem. However, Koschat and Swayne [7] (See also [4]) have proposed an effective iterative algorithm, along with some conjectures about its behavior. We describe a version of it here, establish some of its properties, and report on speed experiments with realistic data for our application. The method is an example of a class of optimization methods called "majorization". See [9] and [6] for more on this method in this and other contexts.

7.1 The general majorization approach

Majorization is an approach to minimizing a real-valued function $f(x)$ as x ranges over some domain M in R^n. To accomplish this, one looks for a majorizing function $g(x, y)$ with the following properties:

1. $f(x) = g(x; x)$ for all x
2. $f(x) \leq g(x; y)$ for all x, y, and
3. for any fixed y, the function $g(x; y)$ is easy to minimize in x.

If f and $g(\cdot; y)$ are smooth for every y, then the first two properties mean that the graphs of f and $g(\cdot; y)$ are tangent at y.

The majorization algorithm then proceeds as follows: start at some x^0. For $k = 0, 1, 2, 3, \dots$, suppose we have defined x^k. Then x^{k+1} is the argument minimizing $g(\cdot; x^k)$. It is easy to see that $f(x^{k+1}) \leq f(x^k)$ for each k.

If M is a compact manifold and $g(\cdot; y)$ has a unique global minimum for each y, then for any starting point x^0, $\{x^i\}$ will converge to a critical point of f, which generically will be a local minimum. For any particular problem, the difficulty is reduced to finding a good majorizing function g, which we now describe for our problem.

7.2 A mapping T on $\mathrm{O}(n)$

Let A, B, C now denote arbitrary real matrices, with A square $n \times n$, B, C rectangular, with compatible sizes so the expressions below make sense. We now drop the subscript F on the Frobenius norm $||.||_F$. Our optimization problem is equivalent to a minimizing a function of the form

$$f(O) = ||AOB - C||^2 \qquad (6)$$

as O ranges over the orthogonal group. The Koschat-Swayne idea is to examine the augmented matrices

$$\left\| \begin{pmatrix} A \\ A_r \end{pmatrix} OB - \begin{pmatrix} C \\ C^* \end{pmatrix} \right\|^2 \qquad (7)$$

where A_r and C^* will be specified later.

This is equal to

$$||AOB - C||^2 + ||A_r OB - C^*||^2,$$

and, expanding via the trace, the terms quadratic in O are

$$tr(AOBB^T O^T A^T) + tr(A_r OBB^T O^T A_r^T)$$

$$= tr(B^T O^T (A^T A + A_r^T A_r) OB).$$

The quadratic dependence on O will drop out if $A^T A + A_r^T A_r = rI$, for some scalar r. If so, the problem can then be solved by the usual orthogonal procrustes method described in Section 6.

So for any r larger than the square of the largest eigenvalue of A, we let

$$A_r^2 = rI - A^2,$$

and then we can minimize, in terms of one singular value decomposition, the objective

$$\left\| \begin{pmatrix} A \\ A_r \end{pmatrix} OB - \begin{pmatrix} C \\ C^* \end{pmatrix} \right\|^2$$

for any fixed choice of C^*, which is still at our disposal. Note that the minimizing O is a unique global minimum when the matrices involved have maximum rank, which we assume from now on.

The method of Koschat and Swayne is to fix some r as above, choose $O_0 \in \mathrm{O}(n)$ at random, and define the sequence $\{O_i\}$ in $\mathrm{O}(n)$ such that O_{i+1} is the unique minimizer of (7) when $C^* = A_r O_i B$.

We can express this in terms of a mapping T as follows.

Definition 7.1 *Choose r as above. Define $T : \mathrm{O}(n) \to \mathrm{O}(n)$ by*

$$T(O) = \mathrm{argmin}_Q(||AQB - C||^2 + ||A_r QB - A_r OB||^2).$$

That is, for any $O \in \mathrm{O}(n)$, $T(O)$ is the unique global minimizer of the function $g(\cdot; O) : \mathrm{O}(n) \to R$ defined by

$$g(Q; O) = ||AQB - C||^2 + ||A_r QB - A_r OB||^2.$$

This g is a majorizing function for f, and the mapping T describes the iteration determined by the majorization algorithm in this case.

Using the method described in Section 6, it is straightforward to justify the following formula for T.

Proposition 7.2 *For T as defined above, and $O \in \mathrm{O}(n)$, let UDV^T be the singular value decomposition of*

$$B(C^T + B^T O^T (rI - A^2)).$$

Then

$$T(O) = VU^T.$$

7.3 Properties of T

Proposition 7.3 *T decreases the objective (6). Away from fixed points, T strictly decreases the objective.*

Proof: For any O,

$$||AT(O)B - C||^2 \leq ||AT(O)B - C||^2 + ||A_r T(O)B - A_r OB||^2$$

$$\leq ||AOB - C||^2 + ||A_r OB - A_r OB||^2 = ||AOB - C||^2,$$

with the inequality strict when $T(O) \neq O$. ∎

The majorization iteration is then equivalent to iteration of this mapping T until the objective no longer decreases by more than a preselected tolerance. In fact, the sequence must converge in $\mathrm{O}(n)$.

Proposition 7.4 *For any $O \in \mathrm{O}(n)$, the sequence $\{T^i(O)\}$ converges to a limit in $\mathrm{O}(n)$.*

Proof: As in the proof of Lemma 7.3, we have

$$||AT(O)B-C||^2+||A_rT(O)B-A_rOB||^2 \leq ||AOB-C||^2,$$

and hence

$$||A_rT(O)B-A_rOB||^2 \leq ||AOB-C||^2-||AT(O)B-C||^2.$$

Therefore, for all n,

$$\begin{aligned} & \sum_{i=0}^{n} ||A_rT^{i+1}(O)B - A_rT^i(O)B||^2 \\ \leq & \sum_{i=0}^{n}(||AT^i(O)B - C||^2 - ||AT^{i+1}(O)B - C||^2) \\ = & ||AOB - C||^2 - ||AT^{n+1}(O)B - C||^2 \\ \leq & ||AOB - C||^2. \end{aligned}$$

Therefore the infinite sum is convergent. Under our assumption that A_r and B have full rank, this implies also

$$\sum_{i=0}^{\infty} ||T^{i+1}(O) - T^i(O)||^2 < \infty. \tag{8}$$

By compactness, the sequence $\{T^i(O)\}$ must have a limit point $O^* \in \mathrm{O}(n)$. By (8) and the triangle inequality, this limit point must be unique; hence O^* must be the limit of the convergent sequence $\{T^i(O)\}$. ∎

The mapping T is not continuous because the singular value decomposition is not continuous in the data. However, it is continuous except on a set of codimension 1, and hence almost everywhere. Thus, with probability one, the limit O^* will be a point of continuity of T, in which case it must then be a fixed point of T.

Proposition 7.5 *Any fixed point Q^* of T is a critical point of the objective $f(O) = ||AOB - C||^2$.*

Proof: Let Q^* be a fixed point of T. With the previous notation

$$g(Q;O) = ||AQB - C||^2 + ||A_rQB - A_rOB||^2,$$

this means that $g(Q;Q^*)$ is minimized by $Q = Q*$.

Let M be a skew-symmetric matrix representing a tangent vector to $\mathrm{O}(n)$, and let $Q_t = Q^* \exp(I + tM)$ represent a path in $\mathrm{O}(n)$ through Q^* in the direction M.

Then

$$g(Q_t;Q^*) \geq g(Q^*;Q^*),$$

or

$$||AQ_tB-C||^2+||A_rQ_tB-A_rQ^*B||^2 \geq ||AQ^*B-C||^2.$$

This means

$$||AQ_tB-C||^2-||AQ^*B-C||^2 \geq -||A_rQ_tB-A_rQ^*B||^2$$

or

$$f(Q_t) - f(Q^*) \geq -||A_rQ^*(\exp(I + tM) - I)B||^2.$$

Since, $\exp(I+tM) = I+tM+O(t^2)$, for small t, the right hand side is of order t^2, and this means all directional derivatives of f at Q^* are non-negative. Hence they must be zero and Q^* must be a critical point of f.

∎

The limiting fixed point is not necessarily the global minimum. If, as we tend to observe, the sequence does not land on the fixed point limit in a finite number of steps, then this fixed point cannot be a local maximum, and so will be either a saddle point, or, generically, a local minimum. Different starting values of O can be expected to lead to different local minima of the objective. Our approach is then to choose a collection of different starting values, perhaps at random, iterate to local minima of the objective, and then choose the smallest of the minima found. Experimentally, we tend to find, with our data, that the different starting values usually lead to the same or similar objective values. Therefore we need not be too concerned with the starting values, since the global minimum has no special advantage for the problem.

7.4 Summary of numerical experiments

We looked at covariance data coming from actual equity risk factors, as described in [1], taken from the MSCIBarra equity risk model as of April 2001. The largest problem corresponds to the 65×65 US equity block, in a 730×730 global covariance matrix. This corresponds to $n = 65$ and B, C of size 65×665, and an optimization over $\mathrm{O}(65)$ of dimension 2080. A simple MATLAB implementation of the iteration on an inexpensive laptop (circa 2005) took about 5 minutes for each starting value of O. For the 27×27 Singapore equity block, the process converged to within 2e-6 after 2200 iterations in about 42 seconds.

By comparison, with the previous Levenberg-Marquardt implementation in C++ on a unix workstation (circa 2000), 20×20 blocks took over an hour and the 65×65 problem did not converge before the experimenters gave up after several hours.

Clearly the Koschat-Swayne approach more efficiently takes advantage of the special algebraic structure of our high-dimensional nonconvex problem, and is good enough for commercial implementation.

8 Conclusion

We have developed a numerically efficient way to adjust a multimarket covariance matrix forecast by correcting one or more diagonal blocks without loss of positive semidefiniteness, and to do so in a way that does not disturb the factors in the other blocks and approximately minimizes the disturbance to the cross-block covariances. In practice, this will allow a global risk forecast to be made consistent with potentially more refined individual market forecasts corresponding to diagonal blocks of the global covariance matrix.

There remains a need for better theoretical understanding of the structure of our objective function. The proposed method still only finds local minima, though in practice we observe that the lowest local minima have the largest basins, and so are the most likely to be found with this method. We feel the special structure of the problem has not yet been fully exploited, leaving open the possibility of a fast method that guarantees a fast approach to the global minimum.

Progress on the Double Orthogonal Procrustes Problem would be of interest in a variety of fields where this data-fitting context arises.

Acknowledgement: This paper continues the work begun in [1], which contains more discussion of the problem. Conversations with G. Anderson, K. Gallivan, L. Goldberg, E. Klassen, R. Tütüncü and J. Zhang are gratefully acknowledged. All errors and omissions are my own.

References

[1] G. Anderson, L. Goldberg, A. Kercheval, G. Miller, and K. Sorge, *On the aggregation of local risk models for global risk management*, Journal of Risk **8** (2005), no. 1, 25–40.

[2] G. Anderson and A. Kercheval, *Correcting negative eigenvalues for estimated correlation matrices in risk management*, Gainesville, FL, April, 2005. Presented at the RMFE conference on Risk Management and Quantitative Approaches to Finance.

[3] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 2nd edition, Johns Hopkins Univ. Press, Baltimore, 1989.

[4] J. Gower and G. Dijksterhuis, *Procrustes Problems*, Oxford Univ. Press, Oxford, UK, 2004.

[5] N. Higham, *Computing the nearest correlation matrix – a problem from finance*, IMA Journal of Numerical Analysis **22** (2002), 329–343.

[6] H. A. L. Kiers and J. M. F. ten Berge, *Minimization of a class of matrix trace functions by means of refined majorization*, Psychometrika **57** (1992), 371–382.

[7] M.A. Koschat and D.F. Swayne, *A weighted procrustes criteria*, Psychometrika **56** (1991), 229–239.

[8] J. Malick, *A dual approach to semidefinite least-squares problems*, SIAM J. Matrix Anal. Appl. **26** (2004), 272–284.

[9] R. Pietersz and P. J. F. Groenen, *Rank reduction of correlation matrices by majorization*, Quantitative Finance **4** (2004), 649–662.

[10] R. Rebonato and P. Jäckel, *The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes*, Journal of Risk **2** (2000), no. 2, 17–27.