

ON NEAR(EST) CORRELATION MATRIX

PASHA ZUSMANOVICH

ABSTRACT. We present an elementary heuristic reasoning based on Arnold's theory of versal deformations in support of a straightforward algorithm for finding a correlation matrix near the given symmetric one.

INTRODUCTION

Bankers are interested in correlations between time series associated with various financial instruments (such as prices of stocks, options, futures and other derivatives, currency exchange rates, etc.), presented in the form of a sample correlation matrix. As a *bona fide* correlation matrix, it should be positive semidefinite. In practice, however, the computed matrix almost always turns out to be not positive semidefinite. The main reason for this is twofold: methodological errors (taking data for different instruments in different time ranges, inconsistent approach to inventing of missing data), and floating point rounding errors.

The computed correlation matrix is utilized, however, in further analysis, like evaluation of various risks; for this, its positive semidefiniteness is crucial. As in the most cases it is impossible to backtrack the origin of the problem due to shortage of time, the large amount of numerical data (a typical scenario may involve daily computed correlation matrices reaching the size of ten thousands by ten thousands), and complexity of the methods used in its retrieval, processing and storage, one usually resorts on “correcting” the symmetric matrix at hand to make it positive semidefinite.

Naturally, this “correction” should be as small as possible. So, a practical problem arises: for a given symmetric matrix, find the nearest, in some sense, correlation matrix. A quick glance at the literature (mentioned below) suggests that this problem arises not only in banking.

Not surprising then that this problem attracted a considerable attention. While the exact expression for the nearest correlation matrix is not available, many papers – see [BH], [QXX] and references therein – contain algorithms for its determination. These algorithms utilize methods from convex analysis, semismooth optimization, and other sophisticated branches of numerical mathematics. Earlier results in this direction are also surveyed in [Ge, §9.4.6]. In all these works, “nearest” is understood in the sense of the Frobenius matrix norm, or some its (weighted) variation.

In the real life, however, bankers tend to ignore all this wisdom and implement a very pedestrian approach to this problem (sometimes called “shrinking” and which can be found, with some variations, in [DI], [QXX], [RM], [Ge, Exercise 9.14], and in many other places). Namely, in the spectral decomposition $A = BJB^T$ of a given $n \times n$ symmetric matrix A , where B is an orthogonal matrix of eigenvectors, and

$$(1) \quad J = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

Date: First written January 26, 2012. Last revised August 5, 2012.

2010 Mathematics Subject Classification. 14D99, 15A18, 15A21, 15B48, 62P05, 62P35, 65F35.

Key words and phrases. Correlation matrix; positive semidefinite matrix; matrix nearness problem; versal deformations of matrices.

is a diagonal matrix of eigenvalues of A , replace all negative eigenvalues by some small positive number ε :

$$\widehat{\lambda}_i = \begin{cases} \varepsilon & \text{if } \lambda_i < 0 \\ \lambda_i & \text{if } \lambda_i > 0, \end{cases}$$

for $i = 1, \dots, n$ (in practice, zero eigenvalues do not occur). The resulting matrix

$$(\widehat{a}_{ij})_{i,j=1}^n = B \begin{pmatrix} \widehat{\lambda}_1 & & & \\ & \widehat{\lambda}_2 & & \\ & & \ddots & \\ & & & \widehat{\lambda}_n \end{pmatrix} B^\top$$

is a positive definite covariance matrix, and its normalization

$$(2) \quad \left(\frac{\widehat{a}_{ij}}{\sqrt{\widehat{a}_{ii}\widehat{a}_{jj}}} \right)_{i,j=1}^n$$

is declared to be the requested correlation matrix, allegedly close to the initial matrix A .

This pedestrian approach turns out to be very efficient in practice (in all banking numerical examples we have observed, the initial and corrected matrices were very close with respect to the max norm[†], and no discrepancies occurred utilizing the corrected matrix in the subsequent analysis). In this note we offer a heuristic argument explaining this, perhaps, unreasonable at the first glance, efficiency. The argument, presented in §2, is an easy application of Arnold's theory of versal deformation of matrices. A fragment of the theory needed for our purposes is briefly recalled in §1.

1. ARNOLD'S THEORY OF VERSAL DEFORMATIONS

In 1971, Vladimir Arnold developed a theory of versal deformations of matrices, which triggered a wave of subsequent work. The original paper [A] is still the best exposition of this theory. The main result of this theory can be formulated in many different ways, one of them runs as follows.

Let A be a complex $n \times n$ matrix with distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$, and with a Jordan normal form

$$J = \begin{pmatrix} J_{n_{11}, \dots, n_{1m_1}}^{\lambda_1} & & & \\ & J_{n_{21}, \dots, n_{2m_2}}^{\lambda_2} & & \\ & & \ddots & \\ & & & J_{n_{k1}, \dots, n_{km_k}}^{\lambda_k} \end{pmatrix}$$

where

$$J_\ell^\lambda = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda & 1 \\ & 0 & & & \lambda \end{pmatrix}$$

[†] In §2, we use at a certain place submultiplicativity (i.e., $\|AB\| \leq \|A\|\|B\|$ for any two matrices A, B) of the matrix norm $\|\cdot\|$ measuring the "nearness". The max norm is not submultiplicative, so, formally, it does not fit those arguments. This can be remedied, however, by a minor (and well-known) fix: the max norm becomes submultiplicative when multiplied by the matrix size (see, for example, [HJ, p. 292]). Even taking into account this factor ($< 10^5$ in practice), the absolute values of differences between the corresponding elements of the initial and corrected matrices remained very small in all examples we have seen.

If the spectrum of A is simple, than this picture is significantly streamlined. The total number N of parameterizing functions in (3) is equal to n , the size of the matrix, and the deformation family of the diagonal matrix (1) itself consists of diagonal matrices:

$$J(\xi_1(\mathbf{t}), \dots, \xi_n(\mathbf{t})) = \begin{pmatrix} \lambda_1 + \xi_1(\mathbf{t}) & & & \\ & \lambda_2 + \xi_2(\mathbf{t}) & & \\ & & \ddots & \\ 0 & & & \lambda_n + \xi_n(\mathbf{t}) \end{pmatrix}$$

where $\xi_i(\mathbf{0}) = 0$, $i = 1, \dots, n$.

There are corresponding results for matrices with real coefficients [Ga] and symmetric matrices [PR] (as well as for many other situations in which a Lie group acts on a manifold, see [S]), which are technically more complicated. However, for our purpose it suffices to use Arnold's original setting. Just note that as we are interested solely in symmetric matrices which are brought to the diagonal form (1) by an orthogonal transformation, the combination of results of [A] and [PR] shows that in the decomposition (3) we may assume that all matrices in the family B are orthogonal, i.e.

$$B(\xi_1(\mathbf{t}), \dots, \xi_n(\mathbf{t}))^{-1} = B(\xi_1(\mathbf{t}), \dots, \xi_n(\mathbf{t}))^\top$$

for all \mathbf{t} from an appropriate neighborhood of zero.

2. JUST GETTING RID OF NEGATIVE EIGENVALUES IS ENOUGH

It is well-known that the set of the correlation matrices coincides with the set of (real) positive semidefinite matrices with units on the main diagonal (see, for example, [F, Chapter III, §6, Theorem 4]), so we will use that two notions interchangeably.

Suppose A is a symmetric $n \times n$ matrix with units on the main diagonal. As the set of matrices with a simple spectrum is Zariski-dense in the set of all real $n \times n$ matrices, we may assume that A has a simple spectrum (a more down-to-earth incarnation of this fact is that all correlation matrices appearing in banking practice, and, more generally, correlation matrices based on a sufficiently large amount of real-world data, have simple spectrum; in fact, the reasonings below could be modified for the case of arbitrary spectrum, but technically they would become more complicated). Let (1) be its Jordan normal form, all λ_i 's being pairwise distinct (and some of them are negative, of small absolute value).

Suppose further that there exists a correlation matrix C "near" A , and that A and C are members of a smooth family of matrices. The latter assumption is justified both from theoretical (correlation matrix is a smooth function of time series it correlates between) and practical (the financial processes a correlation matrix is trying to capture, are assumed to be satisfactorily modelled by smooth functions) perspectives.

According to the theory presented in §1, in a sufficiently small neighborhood \mathcal{U} of $A = A(\mathbf{0})$, we may write this smooth family in the following parametric form:

$$(4) \quad A(\mathbf{t}) = B(\mathbf{t}) \begin{pmatrix} \lambda_1 + \xi_1(\mathbf{t}) & & & \\ & \lambda_2 + \xi_2(\mathbf{t}) & & \\ & & \ddots & \\ 0 & & & \lambda_n + \xi_n(\mathbf{t}) \end{pmatrix} B(\mathbf{t})^\top$$

for some smooth functions ξ_i such that $\xi_i(\mathbf{0}) = 0$ for all $i = 1, \dots, n$, and a smooth family $B(\mathbf{t}) = (b_{ij}(\mathbf{t}))_{i,j=1}^n$ of orthogonal matrices. In particular, C , being a member of the family, is represented in the form (4) for some value $\mathbf{t} = \mathbf{t}_0$.

The condition of positive definiteness of a member of the family $A(\mathbf{t})$ is equivalent to

$$(5) \quad \xi_i(\mathbf{t}) > -\lambda_i$$

for all $i = 1, \dots, n$, and the condition of having units on the main diagonal is equivalent to

$$(6) \quad B^{\circ 2}(\mathbf{t}) \begin{pmatrix} \lambda_1 + \xi_1(\mathbf{t}) \\ \lambda_2 + \xi_2(\mathbf{t}) \\ \vdots \\ \lambda_n + \xi_n(\mathbf{t}) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix},$$

where $B^{\circ 2}(\mathbf{t}) = (b_{ij}(\mathbf{t})^2)_{i,j=1}^n$ is the Hadamard square of $B(\mathbf{t})$. The set of solutions of (6) is nonempty (it contains at least two points, $\mathbf{0}$ and \mathbf{t}_0), hence it forms a hypersurface \mathcal{H} in the space of parameters, and the intersection of this hypersurface with the neighborhood \mathcal{U} and the open domain defined by conditions (5), defines a certain neighborhood \mathcal{U}' of $C = A(\mathbf{t}_0)$ in \mathcal{H} .

In terms of the procedure described in the introduction, getting C from A amounts to “adjusting” eigenvalues, i.e., adding to each eigenvalue λ_i in the diagonal form (1) a small correction $\xi_i(\mathbf{t}_0)$, and subsequent “normalization” (2), what corresponds to getting back correlation matrix

$$B(\mathbf{t}_0) \begin{pmatrix} \lambda_1 + \xi_1(\mathbf{t}_0) & & & \\ & \lambda_2 + \xi_2(\mathbf{t}_0) & & \\ & & \ddots & \\ & & & \lambda_n + \xi_n(\mathbf{t}_0) \end{pmatrix} B(\mathbf{t}_0)^\top$$

from the covariance matrix

$$B(\mathbf{0}) \begin{pmatrix} \lambda_1 + \xi_1(\mathbf{t}_0) & & & \\ & \lambda_2 + \xi_2(\mathbf{t}_0) & & \\ & & \ddots & \\ & & & \lambda_n + \xi_n(\mathbf{t}_0) \end{pmatrix} B(\mathbf{0})^\top.$$

Assuming that the neighborhood \mathcal{U}' is small enough, any matrix from it will do, but what will be the best choice? As mentioned in the introduction, this is, generally, a difficult problem not admitting a closed-form solution. Intuitively, there is no need to adjust the positive eigenvalues, but only the negative ones, and the following imprecise reasoning supports this.

Assuming that the matrix norm $\|\cdot\|$ measuring the “nearness” is submultiplicative and is invariant under transposition (the latter assumption is not essential but slightly simplifies the expressions below), we have:

$$(7) \quad \begin{aligned} & \|A(\mathbf{t}) - A(\mathbf{0})\| \\ &= \|B(\mathbf{t}) \operatorname{diag}(\lambda_1 + \xi_1(\mathbf{t}), \dots, \lambda_n + \xi_n(\mathbf{t})) B(\mathbf{t})^\top - B(\mathbf{0}) \operatorname{diag}(\lambda_1, \dots, \lambda_n) B(\mathbf{0})^\top\| \\ &\leq \|\operatorname{diag}(\xi_1(\mathbf{t}), \dots, \xi_n(\mathbf{t}))\| \|B(\mathbf{t}) - B(\mathbf{0})\|^2 \\ &+ 2\|\operatorname{diag}(\xi_1(\mathbf{t}), \dots, \xi_n(\mathbf{t}))\| \|B(\mathbf{t}) - B(\mathbf{0})\| \|B(\mathbf{0})\| \\ &+ \|\operatorname{diag}(\xi_1(\mathbf{t}), \dots, \xi_n(\mathbf{t}))\| \|B(\mathbf{0})\|^2 \\ &+ \|B(\mathbf{t}) - B(\mathbf{0})\|^2 \|\operatorname{diag}(\lambda_1, \dots, \lambda_n)\| \\ &+ 2\|B(\mathbf{t}) - B(\mathbf{0})\| \|B(\mathbf{0})\| \|\operatorname{diag}(\lambda_1, \dots, \lambda_n)\|. \end{aligned}$$

Both theoretical considerations in [A], and computational procedures developed in [M] suggest that matrix entries of the parametric family $B(\mathbf{t})$ providing the transformation to the canonical form (4) of the versal deformation, have, as power series of the parameter \mathbf{t} , the same order of magnitude as matrix entries of the canonical form itself. In particular, in a sufficiently small neighborhood of zero, which can be assumed lying inside \mathcal{U} ,

$$\|B(\mathbf{t}) - B(\mathbf{0})\| \leq \alpha \|\operatorname{diag}(\xi_1(\mathbf{t}), \dots, \xi_n(\mathbf{t}))\|$$

for some (positive) constant α . This, together with (7), implies that $\|A(\mathbf{t}) - A(\mathbf{0})\|$ is bounded by a cubic polynomial in $\|\operatorname{diag}(\xi_1(\mathbf{t}), \dots, \xi_n(\mathbf{t}))\|$ with positive coefficients. The latter polynomial is a monotonic function, so to minimize $\|A(\mathbf{t}) - A(\mathbf{0})\|$ one may wish to minimize

$\|\text{diag}(\xi_1(\mathbf{t}), \dots, \xi_n(\mathbf{t}))\|$ instead. Subject to restriction (5), for all matrix norms appearing in practice, this amounts to setting $\xi_i(\mathbf{t})$ to a positive value “just a little bit” bigger than $-\lambda_i$ if λ_i is negative, and to zero otherwise.

We stress that these are merely non-rigorous, heuristic, arguments, and by no means they can substitute a rigorous analysis given in [BH], [QXX] and similar papers. However, these arguments perfectly suit the practical nature of the problem: one knows *a priori* that a very close correlation matrix exists. In such a situation, Arnold’s theory guarantees existence of such matrix in the simple form (4). Though it is not guaranteed that this will be the *nearest* correlation matrix, it certainly will be a *near* one, and this suffices in practice.

Of course, arguments of this sort can be used in other similar situations – for example, to justify adjusting (“cutoff”) of some unwanted, from the physical perspective, eigenvalues of (valid) correlation matrices arising in lattice gauge theory (see [YJLL] and references therein), or correcting the degenerate covariance matrix from an insufficient amount of data in the situation when the number of observations is much less than the number of variables (see [TW] and references therein).

ACKNOWLEDGEMENTS

This note was essentially written around 2005, during my employment at ING Bank. Thanks are due to Manon ten Voorde, who, in those days, introduced me to the problem of finding the nearest correlation matrix and explained its peculiarities, and to Dimitry Leites for remarks which improved the presentation. During the final write-up I was supported by grants ERMOS7 (Estonian Science Foundation and Marie Curie Actions) and ETF9038 (Estonian Science Foundation).

REFERENCES

- [A] V.I. Arnold, *On matrices depending on parameters*, Russ. Math. Surv. **26** (1971), $\mathcal{N}2$, 29–43; reprinted in: Singularity Theory. Selected Papers, London Math. Soc. Lect. Note Ser. **53** (1981), 46–60; Russian original reprinted in: Izbrannoe-60, Phasis, 1997, 155–173.
- [BH] R. Borsdorf and N.J. Higham, *A preconditioned Newton algorithm for the nearest correlation matrix*, IMA J. Numer. Anal. **30** (2010), 94–107.
- [DI] J.M. Davenport and R.L. Iman, *An iterative algorithm to produce a positive definite correlation matrix from an “approximate correlation matrix”*, Report SAND-81-1376, Sandia National Laboratories, 1981.
- [F] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd ed., Wiley, 1971.
- [Ga] D.M. Galin, *On real matrices depending on parameters*, Uspekhi Mat. Nauk **27** (1972), $\mathcal{N}1$, 241–242 (in Russian).
- [Ge] J.E. Gentle, *Matrix Algebra. Theory, Computations, and Applications in Statistics*, Springer, 2007.
- [HJ] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge Univ. Press, 1990.
- [M] A.A. Mailybaev, *Transformation to versal deformations of matrices*, Lin. Algebra Appl. **337** (2001), 87–108.
- [PR] J. Patera and C. Rousseau, *Versal deformations of elements of classical Jordan algebras*, J. Math. Phys. **24** (1983), 1375–1380.
- [QXX] H. Qi, Z. Xia and G. Xing, *An application of the nearest correlation matrix on web document classification*, J. Indust. Management Optim. **3** (2007), 701–713.
- [RM] P.J. Rousseeuw and G. Molenberghs, *Transformation of non positive semidefinite correlation matrices*, Comm. Statist. – Theory Meth. **22** (1993), 965–984.
- [S] M.B. Sevryuk, *Comments to problem 1970-1*, Arnold’s Problems (ed. V.I. Arnold), 2nd ed., Springer and Phasis, 2005, 226–231.
- [TW] G.H. Tucci and K. Wang, *New methods for handling singular sample covariance matrices*, arXiv:1111.0235v1 [math.PR].
- [YJLL] B. Yoon, Y.-C. Jang, C. Jung, and W. Lee, *Covariance fitting of highly correlated B_K data*, arXiv:1101.2248v6 [hep-lat].

DEPARTMENT OF MATHEMATICS, TALLINN UNIVERSITY OF TECHNOLOGY, EHITAJATE TEE 5, TALLINN 19086, ESTONIA

E-mail address: pasha.zusmanovich@ttu.ee