Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 30, pp. 592-598, September 2015

ELA

AN IMPROVED ESTIMATE FOR THE CONDITION NUMBER ANOMALY OF UNIVARIATE GAUSSIAN CORRELATION MATRICES*

RALF ZIMMERMANN†

Abstract. In this short note, it is proved that the derivatives of the parametrized univariate Gaussian correlation matrix $R_g(\theta) = \left(\exp(-\theta(x_i - x_j)^2\right)_{i,j} \in \mathbb{R}^{n\times n}$ are rank-deficient in the limit $\theta = 0$ up to any order $m < (n-1)/2$. This result generalizes the rank deficiency theorem for Euclidean distance matrices, which appear as the first-order derivatives of the Gaussian correlation matrices in the limit $\theta = 0$. As a consequence, it is shown that the condition number of $R_g(\theta)$ grows at least as fast as $1/\theta^{\hat{m}+1}$ for $\theta \to 0$, where $\hat{m}$ is the largest integer such that $\hat{m} < (n-1)/2$. This considerably improves the previously known growth rate estimate of $1/\theta^2$ for the so-called Gaussian condition number anomaly.

Key words. Euclidean distance matrix, Gaussian correlation matrix, Almost negative definite matrix, Kriging, Radial basis functions, Vandermonde matrix, Condition number.

AMS subject classifications. 15A57, 15B99, 51K05.

1. Introduction. Let $x_1, \dots, x_n \in \mathbb{R}$ be n mutually distinct points. The univariate *exponential correlation matrix* corresponding to the given point set is defined by

$$R_\alpha(\theta) = (\exp(-\theta(x_i - x_j)^\alpha)_{i,j\leq n} \in \mathbb{R}^{n\times n}, \quad 1 \leq \alpha \leq 2, \tag{1.1}$$

where $\theta \geq 0$ parametrizes the correlation length. The special case of $\alpha = 2$ gives the *Gaussian correlation matrix* $R_g(\theta)$. Exponential correlation matrices appear frequently in spatial statistics and in the design and analysis of computer experiments as well as in radial basis function interpolation (see Koehler and Owen [8], Santner et al. [14], Baxter [2], and Buhmann [4]). In fact, they may be considered as the prototypes of positive definite correlation functions [4, Section 2]. Obviously, $R_\alpha(\theta) \to \mathbf{1}\mathbf{1}^T$ for $\theta \to 0$, where $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^n$ denotes the vector with all entries equal to 1.

In the Gaussian case, it holds $\frac{d}{d\theta}R_g(0) = -\left((x_i - x_j)^2\right)_{i,j\leq n} \in \mathbb{R}^{n\times n}$, that is, up to the sign, the derivative matrix of a Gaussian correlation matrix in $\theta = 0$ is the standard *Euclidean distance matrix (EDM)* corresponding to the point set

*Received by the editors on December 15, 2014. Accepted for publication on September 15, 2015. Handling Editor: Bryan L. Shader.
†Institute of Computational Mathematics, TU Braunschweig, 38100, Germany (ralf.zimmermann@tu-bs.de).

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 30, pp. 592-598, September 2015

ELA

$\{x_1, \dots, x_n\}$. It is a striking fact that the rank of an EDM $E \in \mathbb{R}^{n\times n}$ is independent of n. More precisely, in the univariate case, $\mathrm{rank}(E) \leq 3$, see Lin and Chu [11, Theorem 2.1] for a simple proof. (For more information about EDMs, the interested reader is referred to Gower [6], Alfakih [1], Krislock and Wolkowicz [9], Li et al. [10], and Jaklic and Modic [7].)

The rank deficiency of EDMs has been exploited by Zimmermann [19] in order to establish estimates on the condition number growth rate of $R_\alpha(\theta)$ for $\theta \to 0$, the essential ingredient being the observation that the first-order derivative matrix is rank-deficient in the Gaussian case $\alpha = 2$, while it is invertible in the other exponential cases $1 \leq \alpha < 2$. Hence, the question arises, up to which order m are the derivative matrices of the Gaussian model

$$\frac{d^m}{d\theta^m} R_g(0) = (-1)^m \left((x_i - x_j)^{2m}\right)_{i,j\leq n} \in \mathbb{R}^{n\times n} \tag{1.2}$$

rank-deficient.

In this work, we show that the answer is the largest integer $\hat{m}$ such that $\hat{m} < (n-1)/2$. This fact is used in the further course of the paper to improve the estimate on the extraordinary condition number growth rate (also termed the *condition number anomaly*) of the Gaussian $R_g(\theta)$ for $\theta \to 0$: In [19], it was already indicated that the first-order approach pursued in this reference leads to a lower bound of the growth rate of $\kappa_2(R_g(\theta)) \geq \frac{c}{\theta^2}$ that is far from being sharp. Here, this estimate is improved to a lower bound of $\kappa_2(R_g(\theta)) \geq \frac{c}{\theta^{\hat{m}+1}}$ for $\theta \to 0$.

Since the algebraic proofs of Zimmermann [19] do not readily generalize to exploiting higher-order derivatives, in this work, we pursue a very different approach using tools from linear perturbation theory. This is inspired by a comment of an anonymous referee of [19].

2. Results. First, we settle the question on the rank of the derivative matrices appearing in (1.2).

LEMMA 2.1. *Let $n, p \in \mathbb{N}$ and let $x_1, \dots, x_n \in \mathbb{R}$ be mutually distinct. Let $D_p = ((x_i - x_j)^p)_{i,j\leq n} \in \mathbb{R}^{n\times n}$. Then*

$$\mathit{rank}(D_p) = \min\{p+1, n\}.$$

Proof. It holds

$$D_p = ((x_i - x_j)^p)_{i,j\leq n} = \left(\sum_{k=0}^{p} (-1)^k \begin{pmatrix} p \\ k \end{pmatrix} x_i^{p-k} x_j^k\right)_{i,j\leq n}$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 30, pp. 592-598, September 2015

$$= \sum_{k=0}^{p} (-1)^k \begin{pmatrix} p \\ k \end{pmatrix} \left(x_i^{p-k} x_j^k\right)_{i,j\leq n}$$

$$= \sum_{k=0}^{p} (-1)^k \begin{pmatrix} p \\ k \end{pmatrix} \begin{pmatrix} x_1^{p-k} \\ \vdots \\ x_n^{p-k} \end{pmatrix} \left(x_1^k, \dots, x_n^k\right) \tag{2.1}$$

$$= \left(\mathbf{x}^p, \dots, \mathbf{x}^0\right) \begin{pmatrix} \mu_0 & & \\ & \ddots & \\ & & \mu_p \end{pmatrix} \left(\mathbf{x}^0, \dots, \mathbf{x}^p\right)^T, \tag{2.2}$$

where $\mathbf{x}^k := (x_1^k, \dots, x_n^k)^T \in \mathbb{R}^n$ and $\mu_k := (-1)^k \begin{pmatrix} p \\ k \end{pmatrix}$ for $k = 0, \dots, p$.

The expression (2.1) shows that D_p is the sum of $(p + 1)$ rank-one matrices. As a consequence, $rank(D_p) \leq p + 1$. Morevoer, the vectors $\mathbf{x}^k, k = 0, \dots, p$ are exactly the first $(p + 1)$ columns of the $(n \times n)$–Vandermonde matrix corresponding to the mutually distinct points $x_1, \dots, x_n$. Hence, they are linearly independent, if $p+1 \leq n$. From (2.2), we deduce that in this case $rank(D_p) = p+1$. Hence, $D_p \in \mathbb{R}^{n\times n}$ achieves its maximal rank n for $p = n - 1$ and stays invertible for higher powers of p. ◻

The lemma shows that the matrices $\frac{d^m}{d\theta^m} R_g(0)$ from (1.2) are rank-deficient, if $2m + 1 < n$, and invertible otherwise. The next theorem gives an improved estimate on the Gaussian condition number anomaly.

THEOREM 2.2. *Let $x_1, \dots, x_n \in \mathbb{R}$ be mutually distinct. Let*

$$R_g(\theta) = \left(\exp(-\theta(x_i - x_j)^2\right)_{i,j\leq n} \in \mathbb{R}^{n\times n}$$

be the univariate Gaussian correlation matrix *corresponding to the given point set parametrized by $\theta > 0$. Let $\hat{m}$ be the largest integer such that $\hat{m} < (n - 1)/2$. For $\theta \to 0$, the condition number $\kappa_2(R_g(\theta)$ grows at least as fast as $1/\theta^{\hat{m}+1}$.*

Proof. The Gaussian correlation matrix $R_g(\theta)$ is (strictly) positive definite for all $\theta > 0$, see, e.g., Buhmann [4, Proposition 2.1]. Furthermore, it is analytic in θ. It is a simple exercise to show that for all $m \in \mathbb{N}$, there exist $\varepsilon > 0$ such that the Taylor approximation to order m of $R_g(\theta)$,

$$T_m R_g(\theta) := \sum_{k=0}^{m} \frac{1}{k!} \frac{d^k}{d\theta^k} R_g(0)\theta^k,$$

is positive semi-definite for all $0 \leq \theta \leq \varepsilon$. As has been remarked in the introduction, $\frac{d^m}{d\theta^m} R_g(0) = (-1)^m \left((x_i - x_j)^{2m}\right)_{i,j\leq n}$. By Lemma 2.1,

$$rank\left(\frac{d^m}{d\theta^m} R_g(0)\right) = \min\{2m + 1, n\}$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 30, pp. 592-598, September 2015

and every vector $v \in \left(\text{span}\{\left(x_1^k, \dots, x_n^k\right)^T \mid \quad k = 0, \dots, 2m\}\right)^\perp$ is contained in its nullspace, i.e., $\frac{d^m}{d\theta^m} R_g(0)v = 0$. For such a v, it also holds $\frac{d^k}{d\theta^k} R_g(0)v = 0$ for all $k \leq m$. As a consequence, $T_m R_g(\theta)v = 0$ for such a v, which means that $T_m R_g(\theta)$ has a non-trivial nullspace, whenever $2m + 1 < n$. In summary, if $2m + 1 < n$, then $T_m R_g(\theta)$ is positive semi-definite and singular for all $0 \leq \theta \leq \varepsilon$, so that its smallest eigenvalue is $\lambda_{\min}(T_m R_g(\theta)) = 0$.

Now, let $R_g(\theta) = T_m R_g(\theta) + E(\theta)$ with $E(\theta) = \sum_{k=m+1}^{\infty} \frac{1}{k!} \frac{d^k}{d\theta^k} R_g(0)\theta^k = \mathcal{O}(\theta^{m+1})$. By applying a standard perturbation result for eigenvalues of symmetric matrices (see Golub and Van Loan [5, Corollary 8.1.6]), we obtain

$$|\lambda_{\min}(R_g(\theta)) - \lambda_{\min}(T_m R_g(\theta))| = \lambda_{\min}(R_g(\theta)) \leq \|E(\theta)\|_2 = \mathcal{O}(\theta^{m+1}).$$

Since $\lambda_{\max}(R_g(\theta)) \to n$ for $\theta \to 0$, this proves that for $0 < \theta \to 0$ and any m such that $2m + 1 < n$, it holds

$$\kappa_2(R_g(\theta)) = \frac{\lambda_{\max}(R_g(\theta))}{\lambda_{\min}(R_g(\theta))} \geq \frac{c}{\theta^{m+1}}$$

for some positive constant $c > 0$. $\square$

It is no coincidence that the first-order derivative matrix of the Gaussian correlation model is a Euclidean distance matrix. In fact, this must be so for *every* positive definite correlation model, as shown in Corollary 2.3 below. This is essentially due to the characterization of Euclidean distance matrices as so-called *almost negative matrices* with zero diagonal. By definition, a symmetric matrix $A \in \mathbb{R}^{n\times n}$ is almost negative definite, iff

$$v^T A v \leq 0, \qquad \text{for all } v \in (\text{span}\{\mathbf{1}\})^\perp = \left\{ w \in \mathbb{R}^n \mid \quad \sum_i w_i = 0 \right\}.$$

Almost negative definite matrices $D \in \mathbb{R}^{n\times n}$ with zero diagonal $d_{ii} = 0$, are called distance matrices and are in fact Euclidean. This may be expressed as follows:

$$D \in \mathbb{R}^{n\times n} \text{ is an EDM } \Leftrightarrow \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) D \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \text{ is negative semi-definite.}$$

Note that $(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbb{R}^{n\times n}$ is nothing but the orthogonal projection onto $(\text{span}\{\mathbf{1}\})^\perp$. This result is due to Schoenberg [16] and has been generalized by Gower [6]. It also plays a fundamental role in *multidimensional scaling*, see Mardia et al. [12, Chapter 14]. Various characterizations for almost negative definite matrices are collected by Micchelli [13, Corollary 2.1].

Now, let us consider arbitrary positive semi-definite correlation functions. To this end, let

$$R : [0, \infty) \to \mathbb{R}^{n\times n}, \quad \theta \mapsto R(\theta) \tag{2.3}$$

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 30, pp. 592-598, September 2015

ELA

be an arbitrary real analytic function of positive semi-definite correlation matrices, where the parametrization is chosen without loss of generality such that small values of $\theta > 0$ correspond to strong spatial correlation with the limit case $R(0) = \mathbf{1}\mathbf{1}^T$.

COROLLARY 2.3. The first non-vanishing derivative $-\frac{d^k}{d\theta^k}R(0) \neq 0$ of any real-analytic function of positive definite correlation matrices of the form (2.3) is a Euclidean distance matrix.

Proof. Let $k \in \mathbb{N}$ be the first number such that $\frac{d^k}{d\theta^k}R(0) \neq 0$. Then, the order-$k$ Taylor approximation reduces to $T_k R(\theta) = R(0) + \frac{1}{k!}\frac{d^k}{d\theta^k}R(0)\theta^k$. As argued in the proof of Theorem 2.2, it is positive semi-definite for all $\theta \in [0, \varepsilon]$. In particular

$$0 \leq v^T \left(T_k R(\theta)\right) v = v^T \left(\mathbf{1}\mathbf{1}^T + \frac{1}{k!}\frac{d^k}{d\theta^k}R(0)\theta^k\right) v, \quad \text{for all } v \in (\operatorname{span}\{\mathbf{1}\})^\perp.$$

Hence, $-\frac{1}{k!}\frac{d^k}{d\theta^k}R(0)$ is almost negative definite with zero diagonal and is thus a Euclidean distance matrix by Schoenberg's characterization theorem cited above. ☐

Almost negative definite matrices arise from almost negative kernel functions and are strongly related to positive (semi-)definite kernels. Indeed, a kernel function $r \mapsto \psi(r)$ is almost negative definite if and only if $\exp(-\theta\psi(r))$ is positive definite for all $\theta > 0$ if and only if $\psi(\sqrt{r})$ is completely monotonic, which means that $(-1)^k\psi^{(k)}(\sqrt{r}) \geq 0$ for all $k \in \mathbb{N}_0$. The last equivalence is essentially based on the Bernstein representation of completely monotonic functions (see Schilling et al. [15, Theorem 1.4]) with $\exp(-\theta r)$ as the fundamental building block. The foundations of these results are again due to Schoenberg [17, 18], enhancements as well as textbook proofs may be found in Berg et al. [3, Chapter 3] and Schilling et al. [15, Chapter 4].

Hence, there might be a possibility to derive Corollary 2.3 from the intricate interplay between almost negative and positive definite functions sketched above. However, this seems to be much more involved than the simple proof via the Taylor argument.

Corollary 2.3 has an interesting implication on the dimensionality of the EDMs arising from Gaussian and exponential correlation matrices: In [19], it is exposed that the derivative of the Gaussian correlation matrix in $\theta = 0$ being singular makes the key difference in the condition number growth when compared to the other members of the exponential correlation family

$$(\exp(-\theta|x_i - x_j|^\alpha))_{ij}, \quad 1 \leq \alpha < 2.$$

The condition number growth of this matrix family is bounded by $\mathcal{O}(1/\theta)$ for $\theta \to 0$, see [19, Section 2.2 and Theorem 2]. Yet, Corollary 2.3 reveals the derivatives *in both cases* to be Euclidean distance matrices (see also Baxter [2, Section 2.3] for an alternative proof).

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 30, pp. 592-598, September 2015

ELA

Therefore, the difference in the condition number behavior between the Gaussian model and the other exponential models is better expressed in terms of the *dimensionality* of the respective derivative distance matrices $(|x_i - x_j|^\alpha)_{ij}, 1 \leq \alpha \leq 2$. By definition [6, Section 3], the *dimensionality* of an EDM $D \in \mathbb{R}^{n\times n}$ is the smallest dimension $d \in \mathbb{N}$ of Euclidean space $\mathbb{R}^d$, which contains a point set $\{p^1, \dots, p^n\} \subset \mathbb{R}^d$ that generates the entries of D via Euclidean reaches $\|p^i - p^j\|^2 = D_{ij}$, $i, j \leq n$.

Gower characterized the dimensionality d of an EDM D as the rank of the double-centered matrix $(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)D(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)$, see [6, Theorem 5]. Combining [2, Example 2.3.3], [19, Propositions 1 and 2] and [6, Theorem 6] shows that the dimensionality of the distance matrices

$$(|x_i - x_j|^\alpha)_{ij} \in \mathbb{R}^{n\times n}, \quad 1 \leq \alpha < 2$$

is the *maximum possible number* $d = n-1$, whenever $x_1, \dots, x_n$ are mutually distinct. Hence, one has to search in $\mathbb{R}^{n-1}$ in order to find a point set that reproduces the univariate non-Euclidean distances $|x_i - x_j|^\alpha$ via Euclidean distances. This is in striking contrast to the dimensionality of $(|x_i - x_j|^2))_{ij}$, which trivially is *the minimum possible number* $d = 1$. For a complete coverage of the interplay between rank and dimensionality of EDM's, see Gower [6].

REFERENCES

[1] A.Y. Alfakih. On the eigenvalues of Euclidean distance matrices. *Computational & and Applied Mathematics*, 27(3):237–250, 2008.

[2] B. Baxter. The interpolation Theory of Radial Basis Functions. PhD Thesis, Cambridge, 1992. Available at `http://www.cato.tzo.com/brad/papers/thesis.pdf`.

[3] C. Berg, J.P.R. Christensen, and P. Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.* Graduate Texts in Mathematics, Vol. 100, Springer, New York, 1984.

[4] M.D. Buhmann. *Radial Basis Functions.* Cambridge Monographs on Applied and Computational Mathematics, Vol. 12, Cambridge University Press, Cambridge, 2003.

[5] G.H. Golub and C.F. Van Loan. *Matrix Computations*, third edition. The John Hopkins University Press, Baltimore, 1996.

[6] J.C. Gower. Properties of Euclidean and Non-Euclidean Distance Matrices, *Linear Algebra and its Applications*, 67:81–97, 1985.

[7] G. Jaklic and J. Modic. On Euclidean distance matrices of graphs, *Electronic Journal of Linear Algebra*, 26:574–589, 2013.

[8] J.R. Koehler and A.B. Owen. Computer experiments, In: S. Ghosh and C.R. Rao (editors), *Design and Analysis of Experiments*, Handbook of Statistics, Vol. 13, Elsevier, Amsterdam, 261–308, 1996.

[9] N. Krislock and H. Wolkowicz. Euclidean distance matrices and applications. In: M.F. Anjos and J.B. Lasserre (editors), *Handbook on Semidefinite, Conic and Polynomial Optimization*, International Series in Operations Research and Management Science, Vol. 166, 879–914, 2012.

Electronic Journal of Linear Algebra ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 30, pp. 592-598, September 2015

ELA

[10] C.K. Li, T. Milligan, and M.W. Trosset. Euclidean and circum-Euclidean distance matrices: Characterizations and linear preservers. *Electronic Journal of Linear Algebra*, 20:739–752, 2010.

[11] M.M. Lin and M.T. Chu. On the nonnegative rank of Euclidean distance matrices. *Linear Algebra and its Applications*, 433(3):681–689, 2010.

[12] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Probability and Mathematical Statistics, Academic Press, London, 1979.

[13] C.A. Micchelli. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2(1):11–22, 1986.

[14] T.J. Santner, B.J. Williams, and W.I. Notz. *The Design and Analysis of Computer Experiments*. Springer, New York, 2003.

[15] R.L. Schilling, R. Song, and Z. Vondracek. *Bernstein Functions, Theory and Applications*. De Gruyter, Berlin, 2009.

[16] I.J. Schoenberg. Remarks to Maurice Fréchet's article 'Sur la definition axiomatique d'une classe d'espace distancies vectoriellement applicable sur l'espace de Hilbert'. *Annals of Mathematics*, 36(3):724–732, 1935.

[17] I.J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(4):522–536, 1938.

[18] I.J. Schoenberg. Metric spaces and completely monotone functions. *Annals of Mathematics*, 39(4):811–841, 1938.

[19] R. Zimmermann. On the condition number anomaly of Gaussian correlation matrices. *Linear Algebra and its Applications*, 466(1):512–526, 2015.