FRACTIONAL EUCLIDEAN DISTANCE MATRICES EXTRAPOLATOR FOR SCATTERED DATA

Natalija Pozniak[1], Leonidas Sakalauskas[2]

University of Applied Sciences / Vilniaus kolegija, Saltoniškių str. 58, 08105 Vilnius, Lithuania[1]
Šiauliai University, Vilnius str. 88, 76285 Šiauliai, Lithuania[2]
E-mail: Natalija.Pozniak@gmail.com; Leonidas.Sakalauskas@mii.vu.lt

Introduction

Euclidean distance matrices (EDM) have been receiving increased attention in recently active fields of research. In this paper, non-singularity of EDM, from the point of application to scattered data extrapolation is explored.

Let us consider the data set of K d-dimensional vectors:

$$X = \left(x_1, x_2, ..., x_K \right),$$
(1)

$x_i \in \Re^d$, $1 \le i \le K$. Assume the values of some response function, obtained at these points by physical measurements or computer simulation, are given:

$$Y = \left(y_1, y_2, ..., y_K \right)^T$$
(2)

Denote by $A = \left[\left(\left| x_i - x_j \right| \right)^{\delta} \right]_{i,j=1}^K$ the $K \times K$ matrix of fractional degrees of square Euclidean distances among pairs of vectors of the set X, called the fractional Euclidean distances matrix (FEDM), where $0 \le \delta \le 1$, $\left| x_i - x_j \right| = (x_i - x_j)^T \cdot (x_i - x_j)$, and $E = (1,1,...,1)^T$ is the vector with K components equal to 1. The important case is of usual Euclidean distances, when $\delta = \dfrac{1}{2}$.

Let us define the linear extrapolator $y : \Re^d \to \Re$ by the following way:

$$y(x) = Y^T \cdot u(x),$$
(3)

where weight functions $u : \Re^d \to \Re^K$ satisfy the following properties:

1) $E^T \cdot u(x) = 1$,
2) $u_i(x) = 1$, if $x = x_i$, $1 \le i \le K$.

The well-known approach for extrapolation is given by Shepard method, [6]):

$$y(x) = \begin{cases} \dfrac{Y^T \cdot w(x)}{E^T \cdot w(x)}, & if \ \left| x_i - x \right| \ne 0 \ \ \forall i \\ y_i & , \ if \ \left| x_i - x \right| = 0, \end{cases}$$
(4)

where the weights are chosen by the following way:

$$w(x) = \left(\frac{1}{|x_1 - x|^\delta}, \frac{1}{|x_2 - x|^\delta}, ..., \frac{1}{|x_K - x|^\delta} \right). \tag{5}$$

It is easy to see that Shepard extrapolator satisfies the definition of the linear extrapolator.

Modeling the response function by random Gaussian field, the following extrapolator has been derived, [5]):

$$y(x) = Y^T \cdot A^{-1} \cdot \left(\tau(x) - E \cdot \frac{E^T \cdot A^{-1} \cdot \tau(x) - 1}{E^T \cdot A^{-1} \cdot E} \right), \tag{6}$$

This formula is obtained as the conditional expectation of the random Gaussian field, to which the conditional variance is relating:

$$s^2(x) = d^2 \cdot \left(\tau(x)^T \cdot A^{-1} \cdot \tau(x) - \frac{\left(E^T \cdot A^{-1} \cdot \tau(x) - 1 \right)^2}{E^T \cdot A^{-1} \cdot E} \right), \tag{7}$$

$$d^2 = \frac{1}{K} \cdot \left(Y^T \cdot A^{-1} \cdot Y - \frac{\left(Y^T \cdot A^{-1} \cdot E \right)^2}{E^T \cdot A^{-1} \cdot E} \right), \tag{8}$$

$\tau(x) = \left(|x_1 - x|^\delta, |x_2 - x|^\delta, ..., |x_K - x|^\delta \right)$. Since the quality of both extrapolators depends on smoothing parameter δ it is of the interest to consider extrapolation models that differ from classical cases $\delta = \frac{1}{2}$ or $\delta = 1$. The properties of extrapolator considered depend on the properties of matrix A, which are investigated in the next section.

1. Fractional Euclidean distance matrices

Properties of the extrapolator defined by (6), (7), (8) have not been studied yet. At first, let us study the properties of the matrices with arbitrary degrees $0 \leq \delta < 1$ of square Euclidean distances in more detail. Denote $K \times K$ unit matrix by I, the K - dimensional vector of ones by E. Let us introduce the kernel matrix, [1], [4], [7]):

$$F = -\frac{1}{2} \cdot \left(I - E \cdot s^T \right) \cdot A \cdot \left(I - s \cdot E^T \right), \tag{9}$$

where $s \in \mathfrak{R}^K$, $s^T \cdot E = 1$.

Since the kernel matrices, when $s = \frac{E}{K}$ and $s = (0,0,...,1)$ are studied most often, it is of the interest to study their properties in general case of s. As one can see below, the main property of matrices with arbitrary fractional degrees $0 \leq \delta < 1$ of square Euclidean distances, is that the kernel matrix is positive semi-definite of rank $K - 1$ if no points coincide in the set (1), whereas the rank of square Euclidean distance matrix can be less than $K - 1$ in general [1], [4], [7].

Theorem 1. The kernel matrix F of the matrix $A = \left[\left(\left(x_i - x_j \right)^T \cdot \left(x_i - x_j \right) \right)^\delta \right]_{i,j=1}^K$ is positive semi-definite of rank $K - 1$, here $x_i \in \mathfrak{R}^p$, $x_i \neq x_j$, $i \neq j$, $1 \leq i,j \leq K$, $s^T \cdot E = 1$, $0 \leq \delta < 1$.

Proof. If $\delta = 0$, then the set X can be taken coinciding with vertices of regular K p - dimensional simplex with edges of length 1, and the theorem proposition is easy verified, taking that $A = E \cdot E^T - I$.

As $0 < \delta < 1$, one can derive:

$$r_{ij}^{\delta} = h_{\delta} \cdot \int_0^{\infty} \frac{1 - e^{-\frac{u^2 \cdot r_{ij}}{2}}}{u^{2 \cdot \delta + 1}} du, \text{ where } h_{\delta} = 1 \Bigg/ \int_0^{\infty} \frac{1 - e^{-\frac{u^2}{2}}}{u^{2 \cdot \delta + 1}} du. \tag{10}$$

The next formula follows from Gaussian integral:

$$e^{-\frac{u^2 \cdot r_{ij}}{2}} = e^{-\frac{u^2 \cdot (x_i - x_j)^T (x_i - x_j)}{2}} = (2\pi)^{-\frac{K}{2}} \cdot \int_{\Re^K} e^{i \cdot v \cdot u \cdot (x_i - x_j)} \cdot e^{-\frac{|v|^2}{2}} \cdot dv. \tag{11}$$

Now after some manipulation one can get sure because of (10), (11) the following quadric:

$$\xi^T \cdot F \cdot \xi = -\xi^T \cdot \frac{1}{2}(I - E \cdot s^T) \cdot A \cdot (I - s \cdot E^T) \cdot \xi =$$

$$= -\frac{1}{2} \cdot (\xi^T - (\xi^T \cdot E) \cdot s^T) \cdot A \cdot (\xi - (\xi^T \cdot E) \cdot s) =$$

$$= \frac{-h_{\delta}}{2 \cdot (2\pi)^{\frac{K}{2}}} \cdot \sum_{i=1}^{K} \sum_{j=1}^{K} (\xi_i - (\xi^T \cdot E) \cdot s_i) \cdot (\xi_j - (\xi^T \cdot E) \cdot s_j) \cdot \int_0^{\infty} \left(\int_{\Re^K} \frac{1 - e^{i \cdot u \cdot v \cdot (x_i - x_j)}}{u^{2 \cdot \delta + 1}} \cdot e^{-\frac{|v|^2}{2}} dv \right) du =$$

$$= \frac{h_{\delta}}{2 \cdot (2\pi)^{\frac{K}{2}}} \cdot \int_0^{\infty} \left(\int_{\Re^K} \frac{\left| \sum_{i=1}^{K} (\xi_i - (\xi^T \cdot E) \cdot s_i) \cdot e^{i \cdot u \cdot v \cdot x_i} \right|^2 - ((\xi^T \cdot E) \cdot (1 - (s^T \cdot E)))^2}{u^{2 \cdot \delta + 1}} \cdot e^{-\frac{|v|^2}{2}} dv \right) du \tag{12}$$

Note, under the theorem condition $s^T \cdot E = 1$ the function under integration in (12) is equal to zero for any value of integration vector v only if $\xi = s$. For any other value $\xi \neq s$, the function under integration is positive for certain values of v, and consequently, the integral (12) is positive, so the kernel matrix is positive semi-definite of rank $K - 1\nabla$

Theorem 2. Under the conditions of Theorem 1, the matrix $A = \left[\left((x_i - x_j)^T \cdot (x_i - x_j) \right)^{\delta} \right]_{i,j=1}^{K}$ is nonsingular, namely, $|A| \neq 0$.

Proof: In order to study determinant of matrix A, let us write down the block matrices:

$$A = \begin{bmatrix} \tilde{A} & a \\ a^T & 0 \end{bmatrix}, \quad F = \begin{bmatrix} \tilde{F} & f \\ f^T & v \end{bmatrix}, \quad I - E \cdot s^T = \begin{bmatrix} \tilde{S} & c \\ g^T & r \end{bmatrix}, \tag{13}$$

where $\tilde{A}, \tilde{F}, \tilde{S}$ are $(K-1) \times (K-1)$ matrices, a, f, c, g are $K-1$ dimensional vectors, v, r are scalars. Let us introduce the block matrices:

$$W = \begin{bmatrix} \tilde{F} & w \\ w^T & 0 \end{bmatrix}, \quad \hat{S} = \begin{bmatrix} \tilde{S} & 0 \\ 0^T & 1 \end{bmatrix}, \tag{14}$$

where $w = -\frac{1}{2}\tilde{S} \cdot a$, 0 is $K-1$ dimensional vector of zeros. Assume without loss of generality $s_K \neq 0$. Next,

$$W = \begin{bmatrix} -\dfrac{1}{2}\tilde{S}\cdot\tilde{A}\cdot\tilde{S}^T - \dfrac{1}{2}c\cdot a^T\cdot\tilde{S}^T - \dfrac{1}{2}\tilde{S}\cdot a\cdot g^T & -\dfrac{1}{2}\tilde{S}\cdot a \\[2mm] -\dfrac{1}{2}a\cdot\tilde{S}^T & 0 \end{bmatrix},$$ (15)

Let us note that the vector c consists of elements equal to K^{th} component s_K of the vector s. Thus, the determinant of matrix W remains the same when adding the Kth row multiplied by $-s_K$ to all other columns of this matrix, [2]. Similarly, this determinant does not change when adding the Kth column multiplied by $-s_i$, to i^{th} row, $1 \le i \le K-1$.
Hence,

$$|W| = \left| -\frac{1}{2}\cdot\hat{S}\cdot A\cdot\hat{S}^T \right| = \left(-\frac{1}{2}\right)^K\cdot\left|\tilde{S}\right|^2\cdot|A|.$$ (16)

On the other hand, using the block matrix determinant formula, [2] for (14), it is easy to make sure that

$$|W| = -\left|\tilde{F}\right|\cdot w^T\cdot\tilde{F}^{-1}\cdot w.$$ (17)

It follows from Theorem 1 that $\left|\tilde{F}\right| > 0$ and $w^T\cdot\tilde{F}^{-1}\cdot w > 0$ as well. Thus, $|W| \neq 0$. This and $\left|\tilde{S}\right| \neq 0$ (because of $s_K \neq 0$) implies $|A| \neq 0$. ∇
As follows from **Theorem 2,** the inverse of matrix A exists, hence the extrapolator expressions (6), (7), (8) are correct and can be calculated for any data set (1) consisting of different points and any $0 < \delta < 1$.

2. Computer modeling

In this study, a set of analytic test functions $y_{TF}(x)$ was chosen (Table 1) for the comparison of the developed extrapolator (6) with the Shepard extrapolator (4).

Table 1. *Test functions and domains*

TEST FUNCTION / Mathematical expression	Test domain
Branin $y_{TF}(x) = \left(x_2 - \dfrac{5x_1^2}{4\pi^2} + \dfrac{5x_1}{\pi} - 6\right)^2 + 10\left(1 - \dfrac{1}{8\pi}\right)\cos x_1 + 10$	[-6,6]
Linear $y_{TF}(x) = x_1^2 + x_1\cos(x_1) + x_2\cos(x_2)$	[1,3]
Rosenbrook $y_{TF}(x) = 100\left(x_2 - x_1^2\right)^2 + \left(1 - x_1\right)^2$	[-5,5]

Considered functions are classical test functions usually applied for testing numerical algorithms. The first one is a Branin function (fourth-order polynomials) which shows a dominant second-order trend. This function has an extremely complex and highly non-linear behavior. The second test function is a Linear function composed of polynomials and trigonometric functions, which shows a strong first-order trend. The Rosenbrook function enable us to study extrapolation in data with both the first and second-order trends.

In the test domain for each test function, N=20 (200) samples consisting of K=200 (20) points were uniformly randomly generated and corresponding extrapolation surfaces according to (6) were created. The approxi-

mation surface was compared with test function values in M=200 points, uniformly randomly distributed in the domain.

To compare the accuracy of the model the True Error parameter of the sample point is introduced following to, [3]):

$$TE = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\frac{1}{M} \sum_{j=1}^{M} \left(y(x_j^i) - y_{TF}(x_j^i) \right)^2} \,, \tag{18}$$

where $y(x)$ - extrapolator defined as shown in (6).

$$TE_{Shepard} = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\frac{1}{M} \sum_{j=1}^{M} \left(y_{Shepard}(x_j^i) - y_{TF}(x_j^i) \right)^2} \,, \tag{19}$$

where $y_{Shepard}(x)$ - Shepard extrapolator defined as shown in (4).

Second statistical estimation parameter - the Mean Square Error (MSE) [3], defined:

$$MSE(y_{MSE}) = \frac{1}{2} \sqrt{\frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} s^2(x_j^i)} \,, \tag{20}$$

where $s^2(x)$ - defined as shown in (7). The results for different δ ($\delta = \frac{1}{3}$, $\delta = \frac{1}{2}$, $\delta = \frac{2}{3}$) are summarized in Table 2 and Table 3.

1. A comparison with a sample size 20 (1 sample consists of 200 points).

Table 2. *Error results*

TEST FUNCTION (Number of sample points 200)	$TE(y(x))$	$TE(y_{Shepard}(x))$	δ	$MSE(y_{MSE}(x))$
BRANIN	9.675	118.076	1/3	16.802
	6.153	98.867	1/2	10.22
	4.083	77.873	2/3	6.42
LINEAR	0.073	1.309	1/3	0.145
	0.041	1.054	1/2	0.078
	0.024	0.796	2/3	0.042
ROSENBROOK	$2.921*10^3$	$1.602*10^4$	1/3	$4.424*10^3$
	$2.12*10^3$	$1.437*10^4$	1/2	$3.048*10^3$
	$1.586*10^3$	$1.232*10^4$	2/3	$2.165*10^3$

2. A comparison with a sample size 200 (1 sample consists of 20 points).

Table 3. *Error results*

TEST FUNCTION (Number of sample points 200)	$TE(y(x))$	$TE(y_{Shepard}(x))$	δ	$MSE(y_{MSE}(x))$
BRANIN	58.048	123.501	1/3	56.295
	48.23	109.614	1/2	42.866
	40.92	97.366	2/3	33.528
LINEAR	0.441	1.396	1/3	0.58
	0.306	1.211	1/2	0.397
	0.209	1.047	2/3	0.269
ROSENBROOK	$1.212*10^4$	$1.68*10^4$	1/3	$1.154*10^4$
	$1.12*10^4$	$1.589*10^4$	1/2	$9.978*10^3$
	$1.049*10^4$	$1.511*10^4$	2/3	$8.86*10^3$

Hence, computer modeling enables us to conclude that the extrapolator (6) presents itself the efficient extrapolator of scattered data, because it significantly outperforms the well-known Shepard extrapolator. Quality of extrapolation depends on the parameter δ and its choice should be a subject other research. In its turn, the conditional variance (7) can be taken as a measure of error of extrapolation.

Conclusions

The approach to scattered data extrapolation constructed by FEDM has been developed in the paper. The resulting model is rather simple and depends on a small set of parameters. Results of extrapolator construction by the approach considered for analytically computed surfaces illustrate its applicability for scattered data extrapolation.

The model developed allows us to represent the information obtained from any number of measurements of objective function obtained computing in a computational code or physical experiment and apply for solving practical extrapolation tasks with scattered data in computer graphics, experimental design, etc. At present, the extrapolator is built using only precise information that some measure gave some value. The extrapolator allows prevision (values in the future) or reconstruction of missing data (values in the past). Of course, the model constructed might be generalised for multimodal case and noisy measurements. The best choice of parameter δ and study of matrices with non-Euclidean distances is the subject of future research.

References

1. Gower J. C., 1984, Multivariate analysis: Ordination, multidimensional scaling and allied topics, *Handbook of Applicable Mathematics: Vol. 6. Statistics* (E. H. Lloyd, Ed.), P. 727-781, Wiley, Chichester.
2. Horn R. A., Johnson C. R., 2013, *Matrix Analysis 2nd ed.*, ISBN 978-0-521-54823-6, Cambridge University Press.
3. Kwon H., Choi S., 2015, A trended Kriging model with R2 indicator and application to design optimization, *Aerospace Science and Technology*, Vol. 43, P. 111–125.
4. Schoenberg I. J., 1935, Remarks to Maurice Frechet's article: Sur la definition axiomatique d'une classe d'espaces vectoriels distancies applicables vectoriellement sur l'espace de Hilbert, *Ann. Math.,* Vol. 36, P. 724-732.
5. Sakalauskas L., 2013, Locally Homogeneous and Isotropic Gaussian Fields in Kriging, *Informatica*, Vol. 24, No. 2, P. 253–274.
6. Shepard D., 1968, A Two Dimensional Interpolation Function for Irregularly Spaced Data, *Proc. ACM 23rd Nat'l Conf.*, P. 517-524.
7. Weinberger K.Q, Sha F., Saul L.K., 2004, Learning a Kernel Matrix for Nonlinear Dimensionality Reduction, *Proceedings of the twenty-first international conference on Machine learning (ICML ,04)*, P. 106-113.

Summary

FRACTIONAL EUCLIDEAN DISTANCE MATRICES EXTRAPOLATOR FOR SCATTERED DATA

N. Pozniak, L. Sakalauskas

The paper deals with application of fractional distance matrices to construct the efficient extrapolator of scattered data. The properties of fractional distance matrices are studied in order to develop the linear extrapolator. Study and comparison of developed extrapolator with Shepard extrapolator is performed by computer simulation.

Keywords: Euclidean distance matrices, fractional distance matrices, Shepard extrapolator, linear extrapolator, scattered data**.**

Santrauka

ATSTUMŲ MATRICŲ SU TRUPMENINIAIS LAIPSNIŲ RODIKLIAIS TAIKYMAS DUOMENŲ EKSTRAPOLIAVIMUI

N. Pozniak, L. Sakalauskas

Darbas yra skirtas atstumų matricų, taikomų ekstrapoliavimui, savybių tyrimui bei jų panaudojimui, kuriant efektyviai veikiantį tiesinį išsibarsčiusių duomenų ekstrapoliatorių. Taikant kompiuterinį modeliavimą, naudojant sukurtą ekstrapoliatorių ir žinomą Šepardo ektrapoliatorių, atlikti skaičiavimai ir palyginti rezultatai.

Prasminiai žodžiai: Atstumų matricos su trupmeniniais laipsnių rodikliais, Euklido atstumas, Šepardo ekstrapoliatorius, tiesinis ekstrapoliatorius, išsibarstę duomenys.