JOURNAL OF APPROXIMATION THEORY 68, 74–82 (1992)

Eigenvalues of Euclidean Distance Matrices

KEITH BALL*

Department of Mathematics, Texas A & M University, College Station, Texas 77843, and Trinity College, Cambridge, England

Communicated by Nira Dyn

Received December 5, 1989; revised October 21, 1990

It is shown that if $x_1, ..., x_n$, $n > 1$ are points in $\mathbb{R}^d$ and $\|x_i - x_j\| \geqslant \varepsilon$ whenever $i \neq j$ then the matrix $A = (a_{ij}) = (\|x_i - x_j\|)$ has an inverse whose norm (as an operator on l_2^n) satisfies

$$\|A^{-1}\| \leqslant \frac{c\sqrt{d}}{\varepsilon}$$

for some absolute constant c. The implication for radial basis function interpolation is discussed together with an observation on the norm of the interpolation operator.

INTRODUCTION

In [S], Schoenberg proved the striking result that if $x_1, ..., x_n$ are distinct points in a Hilbert space then the matrix $A = (\|x_i - x_j\|)_{i,j=1}^n$ is invertible. As a consequence, given data $u_1, ..., u_n \in \mathbb{R}$ at distinct points $x_1, ..., x_n \in \mathbb{R}^d$ it is possible to find a function $f: \mathbb{R}^d \to \mathbb{R}$ of the form

$$f(x) = \sum_{j=1}^{n} \alpha_j \|x - x_j\|$$

interpolating the data: i.e., satisfying

$$f(x_i) = u_i, \qquad 1 \leqslant i \leqslant n.$$

Such an interpolation method behaves naturally with respect to translations, rotations, and dilations of the space $\mathbb{R}^d$. This interpolation method, and a modification of it which is discussed in Section 2, is a generalisation of piecewise linear interpolation on the line. On the basis of Schoenberg's

* Partially supported by NSF DMS-8807243.

0021-9045/92 $3.00

method, quite a lot of recent interest has centered on interpolation by functions of the form

$$f(x)=\sum_{1}^{n} \alpha_j g(\|x-x_j\|),$$

where g is some prespecified function: $[0, \infty) \rightarrow \mathbb{R}$. These interpolation methods and their generalisations are referred to as radial basis interpolation. The papers [B], [J], [M], [B-P] and the surveys [D] and [P] provide references to this recent work and comparisons of the advantages of various interpolation schemes.

For the purpose of implementing such interpolation methods it is important that the matrix A should be well-conditioned, in particular that its eigenvalues should be well bounded away from 0. The purpose of this paper is to show that the eigenvalues of A are not too small if the points $x_1, ..., x_n$ are well-separated. Theorem 1 provides a bound independent of n, the number of points, for well-separated sets in $\mathbb{R}^d$, for each fixed dimension d.

THEOREM 1. *There is a constant $c>0$ so that if $x_1, ..., x_n$ $(n>1)$ are points in $\mathbb{R}^d$ with minimum separation ε (i.e., $\|x_i-x_j\| \geqslant \varepsilon$ if $i \neq j$) then the eigenvalues of the matrix $(\|x_i-x_j\|)_{i,j}$ have absolute value no smaller than*

$$\frac{c\varepsilon}{\sqrt{d}}.$$

Section 1 of this paper contains the proof of Theorem 1 and some remarks on the conditioning numbers of the distance matrices involved. Section 2 contains a brief discussion of the interpolation methods related to Schoenberg's and an observation concerning the norm of the interpolation operator in one case.

Theorem 1 was proved in answer to a question of J. Ward: he and F. Narcowich have extended the method described below to estimate matrix norms arising from other radial basis functions [N-W].

THE INVERTIBILITY ESTIMATE

The proof of Theorem 1 uses an extension of Schoenberg's argument involving integral representations of the norm in Euclidean spaces. A function F: $[0, \infty) \rightarrow \mathbb{R}$ will be said to be of negative type on $\mathbb{R}^d$ if for any sequence of points $x_1, ..., x_n \in \mathbb{R}^d$ and scalar sequence $(\lambda_i)_1^n$ with $\sum \lambda_i = 0$,

$$\sum_{i,j} \lambda_i \lambda_j \cdot F(\|x_i-x_j\|) \leqslant 0.$$

(This definition is not quite in agreement with the usual one involving $F(\|x_i - x_j\|^2)$ but is more convenient here. In the language in [D], the functions of negative type are called conditionally negative of order 1.) The main part of Schoenberg's proof consists in finding an integral representation of the norm on $\mathbb{R}^d$ showing that the function $F: r \mapsto r$ is of negative type on $\mathbb{R}^d$ (for all d). The proof of Theorem 1 depends upon the construction of a function F with the following properties:

$$\begin{aligned} &\text{(a)} \quad F \text{ is of negative type on } \mathbb{R}^d \\ &\text{(b)} \quad F(r) = r \qquad \text{if} \quad r \geqslant \varepsilon \\ &\text{(c)} \quad F(0) = \theta > 0. \end{aligned} \tag{1}$$

If such a function F exists, then, given an ε-separated set $x_1, ..., x_n \in \mathbb{R}^d$ and a scalar sequence $(\lambda_i)_1^n$ with $\sum \lambda_i = 0$,

$$\begin{aligned} \theta \sum_1^n \lambda_i^2 + \sum_{i,j} \lambda_i \lambda_j \|x_i - x_j\| &= \theta \sum_1^n \lambda_i^2 + \sum_{i \neq j} \lambda_i \lambda_j \|x_i - x_j\| \\ &= F(0) \sum_1^n \lambda_i^2 + \sum_{i \neq j} \lambda_i \lambda_j F(\|x_i - x_j\|) \\ &= \sum_{i,j} \lambda_i \lambda_j F(\|x_i - x_j\|) \leqslant 0 \end{aligned}$$

so that $\sum_{i,j} \lambda_i \lambda_j \|x_i - x_j\| \leqslant -\theta \sum_i \lambda_i^2$.

This shows that as a quadratic form on the hyperplane $\{(\lambda_i)_1^n : \sum \lambda_i = 0\}$ the matrix $(\|x_i - x_j\|)$ has all eigenvalues at most $-\theta$. The proof is then completed with the following immediate consequence of the Courant–Fisher Theorem. (The lemma is trivial to check directly.)

LEMMA 2. *Suppose $A = (a_{ij})$ is a symmetric $n \times n$ matrix, $n > 1$, with non-negative entries for which*

$$\sum_{i,j} \lambda_i \lambda_j a_{ij} \leqslant -\theta \sum_j \lambda_i^2 \qquad \textit{whenever} \quad \sum \lambda_i = 0$$

(for some $\theta > 0$). Then all eigenvalues of (a_{ij}) have absolute value at least θ.

The functions of negative type on $\mathbb{R}^d$ where characterized by Schoenberg and von Neumann [N-S]. For $d = 1, 2, ...$ let Ω_d: $[0, \infty) \to R$ be the function defined by

$$\Omega_d(\|x\|) = \int_{S^{d-1}} \cos\langle x, \phi \rangle \, d\sigma_{d-1}(\phi),$$

where σ_{d-1} is the rotation invariant probability on the unit sphere $S^{d-1} \subset \mathbb{R}^d$. Thus, if $g_d\colon \mathbb{R} \to \mathbb{R}$ is the function given by

$$g_d(x) = \begin{cases} \alpha_d(1-x^2)^{(d-3)/2} & \text{if} \quad |x| \leqslant 1 \\ 0 & \text{if} \quad |x| > 1, \end{cases}$$

where α_d is chosen so that $\int_{\mathbb{R}} g_d = 1$ then

$$\Omega_d(t) = \hat{g}_d(t) \qquad \text{for} \quad t \geqslant 0.$$

Note that $\Omega_d(0) = 1$ for each d.

THEOREM N-S. *The function $F\colon [0, \infty) \to \mathbb{R}$ is of negative type on $\mathbb{R}^d$ if and only if there is a (positive) measure μ on $[0, \infty)$ such that*

$$F(r) = F(0) + \int_0^\infty (1 - \Omega_d(rt))\, d\mu(t)$$

for every $r \in [0, \infty)$.

Only the easy (if) part of this theorem will actually be needed here. Since no attempt has been made to determine the best possible estimates holding in Theorem 1, only the case of odd dimension d will be dealt with directly. Since any sequence of points in $\mathbb{R}^{d-1}$ can be embedded in $\mathbb{R}^d$, an estimate of the same order in d follows automatically for even dimensions. For d an odd integer, let γ_d denote the distance in $C[-1, 1]$ of the absolute value function from the space of polynomials of degree $d-1$. Then $\gamma_1 = \frac{1}{2}$, $\gamma_3 = \frac{1}{8}$ and according to Bernstein $\gamma_d \sim 0.282.../d$ as $d \to \infty$. The precise result proved below is the following

THEOREM 1′. *Let $(x_i)_1^n$ be points in $\mathbb{R}^d$ with d odd, for which $\|x_i - x_j\| > \varepsilon$ if $i \neq j$. Then all eigenvalues of the matrix $(\|x_i - x_j\|)$ have size at least*

$$\varepsilon \cdot 2^{d-1} \binom{d-1}{(d-1)/2}^{-1} \cdot \gamma_d \sim \varepsilon \frac{0.35...}{\sqrt{d}}.$$

Proof. Assume without loss of generality that $\varepsilon = 1$. By the remarks preceding Lemma 2 it is enough to find a function F with the properties (1a, b, c) with

$$\theta = 2^{d-1} \binom{d-1}{(d-1)/2}^{-1} \gamma_d.$$

By the Hahn–Banach separation theorem there is an (even) signed measure ν on $[-1, 1]$ satisfying

(i) $\int_{-1}^{1} p(x)\,dv(x)=0$ if p is a polynomial of degree $d-1$,

(ii) $\|v\|=1$ (where $\|v\|$ is the total variation of v), and

(iii) $\int_{-1}^{1} |x|\,dv(x)=\gamma_d$.

Define $f\colon \mathbb{R}\to\mathbb{R}$ by the convolution

$$f(y)=\int_{-1}^{1} |y-x|\,dv(x).$$

Condition (i) implies that

(i)′ $f(y)=0$ if $|y|\geqslant 1$ and

(i)″ $\int_{-1}^{1} p(y)\,f(y)\,dy=0$ if p is a polynomial of degree $d-3$.

By condition (ii), if $t\in\mathbb{R}$,

$$t^2|\hat{f}(t)|=2|\hat{v}(t)|=2\left|\int_{-1}^{1}\cos xt\,dv(x)\right| \leqslant 2\|v\|=2$$

and so

(ii)′ $|\hat{f}(t)|/2\leqslant 1/t^2$ for $t>0$.

Finally, (iii) implies that

(iii)′ $(1/2\pi)\int_{\mathbb{R}}\hat{f}(t)\,dt=f(0)=\gamma_d$.

Now, it is easily checked that for $r\geqslant 0$,

$$r=\frac{2^d}{\pi}\binom{d-1}{(d-1)/2}^{-1}\int_0^\infty (1-\Omega_d(rt))\frac{dt}{t^2}. \tag{2}$$

Define a modification of the function $r\mapsto r$ by

$$F(r)=2^{d-1}\binom{d-1}{(d-1)/2}^{-1}\gamma_d +\frac{2^d}{\pi}\binom{d-1}{(d-1)/2}^{-1}\int_0^\infty (1-\Omega_d(rt))\left(\frac{1}{t^2}-\frac{\hat{f}(t)}{2}\right)dt$$

for $r\geqslant 0$.

By (ii)′ and the theorem of Schoenberg and von Neumann, F is of negative type on $\mathbb{R}^d$, so the first of the three properties (1a) holds. Since $\Omega_d(0)=1$,

$$F(0)=2^{d-1}\binom{d-1}{(d-1)/2}^{-1}\gamma_d$$

so the third property (1c) holds with

$$\theta = 2^{d-1}\binom{d-1}{(d-1)/2}^{-1}\gamma_d$$

as required.

Now, by identity (2), and (iii)′,

$$\begin{aligned} F(r) &= 2^{d-1}\binom{d-1}{(d-1)/2}^{-1}\gamma_d + r \\ &\quad -\frac{2^{d-1}}{\pi}\binom{d-1}{(d-1)/2}^{-1}\int_0^\infty \hat{f}(t)\,dt \\ &\quad +\frac{2^{d-1}}{\pi}\binom{d-1}{(d-1)/2}^{-1}\int_0^\infty \Omega_d(rt)\,\hat{f}(t)\,dt \\ &= r+\frac{c}{\pi}\int_0^\infty \Omega_d(rt)\,\hat{f}(t)\,dt \end{aligned}$$

(with the appropriate c) where the convergence of the separate $\hat{f}$ integrals is a consequence of the facts that $|\hat{f}(t)| \leqslant t^{-2}$ for $t>0$ and $\hat{f}$ is bounded. So, to prove property (1b) for F and complete the theorem, it suffices to show that

$$\frac{1}{\pi}\int_0^\infty \Omega_d(rt)\,\hat{f}(t)\,dt = 0 \qquad \text{if} \quad r \geqslant 1.$$

But since $\Omega_d = \hat{g}_d$, Parseval's theorem implies that the integral is

$$\frac{1}{r}\int_{\mathbb{R}} g_d\left(\frac{x}{r}\right) f(x)\,dx = \frac{1}{r}\int_{-r}^{r}\left(1-\frac{x^2}{r^2}\right)^{(d-3)/2} f(x)\,dx.$$

Now if $r \geqslant 1$, then this expression is just

$$\frac{1}{r}\int_{-1}^{1}\left(1-\frac{x^2}{r^2}\right)^{(d-3)/2} f(x)\,dx$$

since f is supported on $[-1, 1]$; condition (i)′. Finally, since $(1-x^2/r^2)^{(d-3)/2}$ is a polynomial in x of degree $d-3$, condition (i)″ shows that the integral is zero. ∎

For $d=1$, the estimate $\varepsilon/2$ of Theorem 1′ is best possible. For $d=3$ the estimate $\varepsilon/4$ is not best possible and it seems highly unlikely that the estimate is exact for any $d>3$. It would be an interesting geometric problem to determine the best possible estimates, at least for $d=2$ and 3, but even the determination of the best estimates possible with the method used here seems to involve delicate number theoretic considerations.

Using Theorem 1 it is easy to obtain an estimate on the condition number $\|A^{-1}\| \cdot \|A\|$ of a matrix $A = (\|x_i - x_j\|)$ defined by an ε-separated set $[x_1, ..., x_n] \subset \mathbb{R}^d$ of diameter at most D. For at most about $(D/\varepsilon)^d$ points can be found in $\mathbb{R}^d$ satisfying the conditions, so $n \lesssim (D/\varepsilon)^d$ and since each entry in A is at most D, $\|A\| \leqslant nD \lesssim D^{d+1}/\varepsilon^d$. By Theorem 1, $\|A^{-1}\| \, \|A\| \lesssim (D/\varepsilon)^{d+1}$ (where the sign $\lesssim$ indicates inequality up to a constant depending upon d).

2. Some Remarks on the Interpolation Procedure

As stated in the introduction, Schoenberg's theorem shows that given distinct points in $\mathbb{R}^d$, $x_1, ..., x_n$, arbitrary data $u_1, ..., u_n$ at the points can be interpolated by a function of the form

$$f(x) = \sum_1^n \alpha_j \|x - x_j\|.$$

But a function of this form is bounded on $\mathbb{R}^d$ if and only if $\sum_1^n \alpha_i = 0$. Therefore several authors have considered instead, functions of the form

$$f(x) = \delta + \sum_1^n \alpha_j \|x - x_j\|, \tag{3}$$

where δ is some constant and $\sum_1^n \alpha_j = 0$. (This was pointed out to me by M. J. D. Powell.) It is clear from the proof of Theorem 1 (or Schoenberg's original proof) that functions of the form (3) can be used to interpolate arbitrary data. For if $(\lambda_i)_1^n$ is chosen so that $\sum_j \lambda_j \|x_i - x_j\| = 1$ for all i, then $\sum_1^n \lambda_i \neq 0$ since otherwise

$$0 = \sum_i \lambda_i = \sum_{i,j} \lambda_i \lambda_j \|x_i - x_j\| < 0.$$

Hence if $(\beta_i)_1^n$ are chosen so that

$$\sum_j \beta_j \|x_i - x_j\| = u_i \qquad \text{for all } i$$

and δ is such that $\sum \beta_i - \delta \sum \lambda_i = 0$, the numbers $\alpha_i = \beta_i - \delta\lambda_i$ satisfy

$$\sum \alpha_i = 0$$

and

$$\sum \alpha_j \|x_i - x_j\| = u_i - \delta \qquad \text{for each } i.$$

It was shown by Buhmann and Powell [B-P] that interpolation by functions (3) for points in a regular grid is highly localised even if the dimension d is larger than 1 (piecewise linear interpolation on the lime being completely local). It is natural to ask how large can be the norm of the interpolation operator $T: l_\infty^n \to L_\infty(\mathbb{R}^d)$ which takes a sequence $(u_i)_1^n$ to its interpolant f, of the form (3). If $d=1$ then the norm is 1. For $d>1$, however, the norm depends heavily on the number of points n, and/or their distribution.

PROPOSITION 3. *For each $\varepsilon \in (0, 1)$, $n \in \mathbb{N}$, there is an ε-separated set of $2n$ points in $\mathbb{R}^2$, with diameter at most n, for which the norm of the interpolation operator T, described above, is at least*

$$\frac{n}{1+\varepsilon \log n}.$$

Proof. Let $(x_i)_1^n$ and $(y_i)_1^n$ be given

$$x_i = \left(-\frac{\varepsilon}{2}, i\right) \in \mathbb{R}^2,$$

$$y_i = \left(\frac{\varepsilon}{2}, i\right) \in \mathbb{R}^2, \qquad 1 \leqslant i \leqslant n.$$

Define $f: \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x) = \sum_1^n \|x - y_j\| - \sum_1^n \|x - x_j\|, \qquad x \in \mathbb{R}^2.$$

Then f is a function of the form (3) interpolating $(f(x_1), ..., f(x_n), f(y_1), ..., f(y_n))$. Now for $1 \leqslant i \leqslant n$,

$$\begin{aligned} f(x_i) &= \sum_{j=1}^n (\|x_i - y_j\| - \|x_i - x_j\|) \\ &= \sum_{j=1}^n ((\varepsilon^2 + (i-j)^2)^{1/2} - |i-j|) \\ &= \sum_{j=1}^{i-1} ((\varepsilon^2 + (i-j)^2)^{1/2} - (i-j)) \\ &\quad + \varepsilon + \sum_{i+1}^n ((\varepsilon^2 + (j-i)^2)^{1/2} - (j-i)) \\ &\leqslant \sum_{j=1}^{i-1} \frac{\varepsilon^2}{2(i-j)} + \varepsilon + \sum_{i+1}^n \frac{\varepsilon^2}{2(j-i)} \\ &\leqslant \varepsilon + \varepsilon^2 \log n \end{aligned}$$

and since $f(x_i) \geqslant 0$, $|f(x_i)| \leqslant \varepsilon + \varepsilon^2 \log n$. By symmetry, the same estimate holds for $|f(y_i)|$, $1 \leqslant i \leqslant n$.

On the other hand, it is easily checked that if $x = (-R, 0)$ then as $R \to \infty$, $f(x) \to \|\sum_1^n y_j - \sum_1^n x_j\| = n\varepsilon$. Hence the interpolation operator T has norm at least

$$\frac{n\varepsilon}{\varepsilon + \varepsilon^2 \log n} = \frac{n}{1 + \varepsilon \log n}. \quad \blacksquare$$

Simple experiments suggest that for an ε-separated set of diameter D, consisting of n points in $\mathbb{R}^d$, the norm of the interpolation operator can be estimated by a multiple of n or a multiple of $(D/\varepsilon)^{d-1}$. Since a uniform collection of such points will have $n \approx (D/\varepsilon)^d$, the latter estimate might be preferable.

References

[B-P] M. D. Buhmann and M. J. D. Powell, Radial basis function interpolation on an infinite regular grid, *in* "Algorithms for Approximation" (J. C. Mason and M. G. Cox, Eds.), Oxford Univ. Press, Oxford, 1989.

[D] N. Dyn, Interpolation and approximation by radial and related functions, *in* "Approximation Theory VI" (C. K. Chui, L. L. Schumaker, and J. D. Ward, Eds.), Academic Press, San Diego, 1989.

[J] I. R. H. Jackson, Convergence properties of radial basis functions, *Constr. Approx.* **4** (1988), 243–264.

[M] C. A. Micchelli, Interpolation of scattered data: Distance matrices and conditionally positive definite functions, *Constr. Approx.* **2** (1986), 11–22.

[N-S] J. von Neumann and I. J. Schoenberg, Fourier integrals and metric geometry, *Trans. Amer. Math. Soc.* **50** (1941), 226–251.

[N-W] F. J. Narcowich and J. D. Ward, Norms of Inverses and Condition Numbers of Matrices Associated with Scattered Data, preprint.

[P] M. J. D. Powell, Radial basis functions for multivariable approximation: A review, *in* "Algorithms for Approximation" (J. C. Mason and M. G. Cox, Eds.), pp. 143–167, Oxford Univ. Press, Oxford, 1987.

[S] I. J. Schoenberg, On certain metric spaces arising from Euclidean space by a change of metric and their imbedding in Hilbert space, *Ann. of Math.* **38** (1937), 787–793.