JOURNAL OF APPROXIMATION THEORY **60**, 133–140 (1990)

# Some Remarks on the Fejér Problem for Lagrange Interpolation in Several Variables

LEN BOS

*Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4*

*Communicated by E. W. Cheney*

Received September 14, 1987; revised July 2, 1988

Suppose that $X \subset \mathbb{R}^m$ is compact. The polynomials of total degree at most $n$, when restricted to $X$, form a certain vector space, $\mathscr{P}_n(X)$ say. Let $N_m(X)$ be the dimension of $\mathscr{P}_n(X)$. Often, when not ambiguous, we will abbreviate $N_n(X)$ to $N$. If $\{p_1, p_2, ..., p_N\}$ is a basis of $\mathscr{P}_n(X)$ and $\{x_1, x_2, ..., x_N\}$ a collection of $N$ points of $X$, the $N \times N$ matrix

$$V_n(x_1, x_2, ..., x_N) := [p_i(x_j)]$$

is known as the corresponding Vandermonde matrix. We will denote its determinant by

$$\mathrm{VDM}(x_1, x_2, ..., x_N) := \det V_n(x_1, x_2, ..., x_N).$$

If $\mathrm{VDM}(x_1, ..., x_N) \neq 0$ we may form the Lagrange polynomials

$$l_i(x) := \frac{\mathrm{VDM}(x_1, ..., x_{i-1}, x, x_{i+1}, ..., x_N)}{\mathrm{VDM}(x_1, ..., x_N)}, \qquad 1 \leqslant i \leqslant N.$$

Each $l_i$ is a polynomial of degree at most $n$ and is easily seen to have the property that $l_i(x_j) = \delta_{ij}$. Further, if $f \in C(X)$, then $p(x) := \sum_{i=1}^N f(x_i)\, l_i(x)$ is the unique polynomial (when restricted to $X$) of degree at most $n$ which interpolates $f$ at the points $x_1, ..., x_N$. $\Lambda(x) := \sum_{i=1}^N |l_i(x)|$ is known as the Lebesgue function of the interpolation. It is not difficult to see that $\max_{x \in X} \Lambda(x)$ is the norm of the projection

$$f \to \sum_{i=1}^N f(x_i)\, l_i(x).$$

Thus $\Lambda(x)$ gives information on the convergence of interpolants and in one variable has been much studied (see, e.g., [7]). In particular, it is usually



0021-9045/90 $3.00

desirable to interpolate at points for which $\Lambda(x)$ is as small as possible. Now, VDM: $X^N \to \mathbb{R}$ is continuous and thus, as $X$ is compact, attains its maximum. Clearly, for these maximal points, $\max_{x \in X} |l_i(x)| = 1$ and hence $\max_{x \in X} \Lambda(x) \leqslant N$. But, in 1932, Fejér [3] proved the remarkable fact that for $X = [-1, 1]$ or $X = S_1$, the unit circle, at the points which maximize the Vandermonde determinant, $\max_{x \in X} \sum_{i=1}^N l_i^2(x) = 1$ and hence, in these cases, $\max_{x \in X} \Lambda(x) \leqslant N^{1/2}$. It is of some interest to know if $\max_{x \in X} \sum_{i=1}^N l_i^2(x) = 1$ for other regions. In particular, we consider $X = S_{m-1} := \{x \in \mathbb{R}^m : \|x\|_2 = 1\}$, the unit sphere, and $X = B_m := \{x \in \mathbb{R}^m : \|x\|_2 \leqslant 1\}$, the unit ball. Surprisingly, the answer to these questions is related to the statistical theory of optimal experimental design and to the theory of tight spherical designs. Much of the material may be found elsewhere. See especially Karlin and Studden [5, Chap. X]. The purpose of this note is to collect the results relevant to interpolation and use them to conclude that, except in exceptional circumstances, $\max_{x \in X} \sum_{i=1}^N l_i^2(x) > 1$ for these two regions.

We proceed as follows. Clearly, the problem of maximizing $\det(1/N)\, V_n V_n^T$ is the same as maximizing $\det V_n$. But $(1/N)\, V_n V_n^T$ has entries $(1/N) \sum_{k=1}^N p_i(x_k)\, p_j(x_k) = \int_X p_i(x)\, p_j(x)\, d\mu$, where $\mu$ is the discrete probability measure with weight $1/N$ at each of the $N$ points $x_i$. In general, for $\mu$ a probability measure on $X$, let $M(\mu)$ denote the $N \times N$ Gram matrix

$$M(\mu) := \left[\int_X p_i(x)\, p_j(x)\, d\mu\right].$$

Clearly, each $M(\mu)$ is positive semi-definite. Notice also, that if card $\operatorname{supp}(\mu) < N$, $\{p_1, ..., p_N\}$ is linearly dependent on $\operatorname{supp}(\mu)$ and hence $M(\mu)$ is singular. We will generalize our problem and maximize $\det M(\mu)$ over *all* probability measures on $X$.

LEMMA 1 [5, p. 323]. *The family of matrices, $M(\mu)$, as $\mu$ ranges over all probability measures, is a compact convex set.*

Hence $\{\det M(\mu) : \mu$ is a probability measure on $X\}$ is an interval and $\max \det M(\mu)$ is attained. Now set $\mathbf{p}(x) := [p_i(x)] \in \mathbb{R}^N$ and, for $\det M(\mu) \neq 0$,

$$d(x; \mu) := \mathbf{p}^T(x)\, M^{-1}(\mu)\, \mathbf{p}(x).$$

The crucial result is the remarkable equivalence theorem of Kiefer and Wolfowitz [6].

THEOREM 2. *Suppose that $\mu^*$ is a probability measure on $X$. Then $\mu^*$ maximizes* $\det M(\mu)$ *if and only if* $\max_{x \in X} d(x; \mu^*) = N$. *Furthermore, all* *such optimal* matrices *are the same (even though the optimal measure will not usually be unique).*

*Proof.* We offer a slightly simpler proof than that given in [6] or [5]. Suppose that $\mu_1$ and $\mu_2$ are two probability measures on $X$. Then for $t \in [0, 1]$, $(1-t)\,\mu_1 + t\mu_2$ is also a probability measure and

$$M((1-t)\,\mu_1 + t\mu_2) = (1-t)\,M(\mu_1) + tM(\mu_2).$$

Now, as these matrices are symmetric, positive semi-definite (see, e.g., [4, p. 314]), there is a non-singular matrix $A$ such that $A^T M(\mu_1)\,A = \operatorname{diag}(a_1, ..., a_N)$ and that $A^T M(\mu_2)\,A = \operatorname{diag}(b_1, ..., b_N)$ are both diagonal and thus

$$\begin{aligned} \det M((1-t)\,\mu_1 + t\mu_2) &= (\det A)^{-2} \det \operatorname{diag}((1-t)\,a_i + tb_i) \\ &\geqslant (\det M(\mu_1))^{1-t} (\det M(\mu_2))^t \end{aligned} \tag{1}$$

with equality iff $a_i = b_i$, $1 \leqslant i \leqslant N$, i.e., $M(\mu_1) = M(\mu_2)$. Hence if $\mu_1$ and $\mu_2$ both maximize $\det M(\mu)$, $M(\mu_1) = M(\mu_2)$.

Further, from (1), we may easily compute that

$$\frac{\partial^2}{\partial t^2} \log \det M((1-t)\,\mu_1 + t\mu_2) = -\sum_{i=1}^{N} \frac{(b_i - a_i)^2}{((1-t)\,a_i + tb_i)^2} \leqslant 0.$$

Hence $\mu_1$ maximizes $\det M(\mu)$ if and only if

$$\frac{\partial}{\partial t} \log \det M((1-t)\,\mu_1 + t\mu_2)|_{t=0} \leqslant 0$$

for all probability measures $\mu_2$. But,

$$\begin{aligned} &\frac{\partial}{\partial t} \log \det M((1-t)\,\mu_1 + t\mu_2) \\ &\quad= \frac{\partial}{\partial t} \operatorname{tr} \log M((1-t)\,\mu_1 + t\mu_2) \\ &\quad= \operatorname{tr} M^{-1}((1-t)\,\mu_1 + t\mu_2) \frac{\partial}{\partial t} M((1-t)\,\mu_1 + t\mu_2) \end{aligned}$$

and so $\mu_1$ is optimal iff

$$\operatorname{tr} M^{-1}(\mu_1)(M(\mu_2) - M(\mu_1)) = \operatorname{tr} M^{-1}(\mu_1)\,M(\mu_2) - N \leqslant 0.$$

Now an easy computation reveals that

$$\operatorname{tr} M^{-1}(\mu_1)\, M(\mu_2) = \int_X \mathbf{p}^T(x)\, M^{-1}(\mu_1)\, \mathbf{p}(x)\, d\mu_2$$

and so we see that $\mu^*$ maximizes $\det M(\mu)$ if and only if

$$\int_X d(x; \mu^*)\, d\mu \leqslant N$$

for all probability measures $\mu$.

If we take $\mu$ to be concentrated at a single point $x \in X$ we have that if $\mu^*$ maximizes $\det M(\mu)$ then $d(x; \mu^*) \leqslant N$. But as $\int_X d(x; \mu)\, d\mu = \operatorname{tr} M^{-1}(\mu)\, M(\mu) = N$, we have also that $N \leqslant \max_{x \in X} d(x; \mu^*)$ and so, in fact, $\max_{x \in X} d(x; \mu^*) = N$.

Conversely, if $\max_{x \in X} d(x; \mu^*) = N$ then for any other probability measure, $\mu$,

$$\int_X d(x; \mu^*)\, d\mu \leqslant N$$

and so $\mu^*$ maximizes $\det M(\mu)$. ∎

An immediate corollary to the above gives us a criterion for when $\max_{x \in X} \sum_{i=1}^N l_i^2(x) = 1$ is realizable.

THEOREM 3. *Suppose that for* $x_1, ..., x_N \in X$, $\mathrm{VDM}(x_1, ..., x_N) \neq 0$. *Then* $\max_{x \in X} \sum_{i=1}^N l_i^2(x) = 1$ *if and only if the discrete measure,* $\mu^*$, *with weight* $1/N$ *at each* $x_i$ *maximizes* $\det M(\mu)$.

*Proof.* The conclusion of Theorem 2 is independent of the basis of polynomials used. Choose $\{l_1(x), ..., l_N(x)\}$ as the basis. Then $\int_X l_i(x)\, l_j(x)\, d\mu^* = (1/N) \sum_{k=1}^N l_i(x_k)\, l_j(x_k) = (1/N)\, \delta_{ij}$ and so $M(\mu^*) = (1/N)\, I$ and $d(x; \mu^*) = N \sum_{i=1}^N l_i^2(x)$. The result follows. ∎

Alternatively, we may express this condition in terms of equally weighted numerical integration formulas.

THEOREM 4. *Suppose that for* $x_1, ..., x_N \in X$, $\mathrm{VDM}(x_1, ..., x_N) \neq 0$ *and that* $\mu^*$ *maximizes* $\det M(\mu)$. *Then* $\max_{x \in X} \sum_{i=1}^N l_i^2(x) = 1$ *if and only if* $\int_X p(x)\, d\mu^* = (1/N) \sum_{i=1}^N p(x_i)$ *for all polynomials,* $p$, *of degree at most* $2n$.

*Proof.* Let $\mu$ be the discrete measure with weight $1/N$ at each $x_i$. It is easy to see that $M(\mu^*) = M(\mu)$ if and only if $\int_X p(x)\, d\mu^* = \int_X p(x)\, d\mu$ for all polynomials, $p$, of degree at most $2n$. The result follows from the uniqueness of the optimal matrix. ∎

Now, for the sphere we actually find an optimal probability measure. Not surprisingly, it turns out to be normalized surface area.

THEOREM 5 (Cf. [5, p. 344] for a more general result). *Let* $X = S_{m-1}$ *and* $d\mu^* = (1/\omega_{m-1})\, d\sigma$, *where* $\omega_{m-1}$ *is the surface area of* $S_{m-1}$ *and* $d\sigma$ *is the surface area differential. Then* $\mu^*$ *maximizes* $\det M(\mu)$.

*Proof.* Suppose that $g \in SO(N)$. For $\mu$ a probability measure on $S_{m-1}$ let $\mu_g$ be the probability measure given by

$$\int_{S_{m-1}} f(x)\, d\mu_g = \int_{S_{m-1}} f(g^{-1}x)\, d\mu.$$

If we let $d\xi(g)$ be the unit Haar measure on $SO(N)$ then $\int_{SO(N)} \mu_g\, d\xi(g)$ is a probability measure invariant under the action of $SO(N)$ and hence must equal $\mu^*$.

Now, as $\{p_i(g^{-1}x)\}$ is also a basis for the polynomials restricted to the sphere, there is a matrix $A_g \in \mathbb{R}^{N \times N}$ such that $\mathbf{p}(g^{-1}x) = A_g \mathbf{p}(x)$. It is easy to see that $A_{g_1 g_2} = A_{g_2} A_{g_1}$ and as $\{g \in SO(N) : g^k = I \text{ for some } k\}$ is dense is $SO(N)$ we have $\det A_g = \pm 1$. Further,

$$\begin{aligned} M(\mu_g) &= \int_X \mathbf{p}(x)\, \mathbf{p}^T(x)\, d\mu_g \\ &= \int_X \mathbf{p}(g^{-1}x)\, \mathbf{p}^T(g^{-1}x)\, d\mu_g \\ &= \int_X A_g \mathbf{p}(x)\, \mathbf{p}^T(x)\, A_g^T\, d\mu_g \\ &= A_g M(\mu)\, A_g^T, \end{aligned}$$

and so $\det M(\mu_g) = \det M(\mu)$. Similarly,

$$\begin{aligned} M(\mu^*) &= \int_X \mathbf{p}(x)\, \mathbf{p}^T(x)\, d\mu^* \\ &= \int_{SO(N)} \int_X \mathbf{p}(x)\, \mathbf{p}^T(x)\, d\mu_g\, d\xi(g) \\ &= \int_{SO(N)} M(\mu_g)\, d\xi(g). \end{aligned}$$

But by (1), $\det M(\mu)$ is concave in $\mu$ and so by Jensen's inequality $\det M(\mu^*) \geqslant \det M(\mu)$. ∎

For $n=1$, $N_n(S_{m-1})=N_n(B_m)=m+1$. If we place one point at each of the vertices of a regular simplex inscribed in $S_{m-1}$ then we must have $x_i \cdot x_j = -1/m$ for $i \neq j$. Hence $l_i(x)=\{m(x\cdot x_i)+1\}/(m+1)$, $1 \leqslant i \leqslant m+1$. One may then easily calculate that $\sum_{i=1}^{m+1} l_i^2(x) = \{1+m\,\|x\|^2\}/(m+1) \leqslant 1$ for $x \in B_m$ and hence also for $x \in S_{m-1}$. Thus these points maximize the Vandermonde determinant over both $S_{m-1}$ and $B_m$. The situation is somewhat more complicated for higher degrees. Since an optimal measure for $S_{m-1}$ is surface area, $\max_{x\in S_{m-1}} \sum_{i=1}^{N} l_i^2(x)=1$ is realizable iff $\int_{S_{m-1}} p(x)\,d\sigma = (\omega_{m-1}/N)\sum_{i=1}^{N} p(x_i)$ for all polynomials of degree at most $2n$. Such a configuration of points is known as a spherical design of strength $2n$ (see, e.g., [2]). As it is not hard to see that a design of strength $2n$ must consist of at least $N_n(S_{m-1})$ points, a design with exactly $N$ points is said to be tight. Bannai and Dammerel [1] have shown that for $n \geqslant 3$ and $m \geqslant 3$ no such design exists. We therefore have

THEOREM 6. *Suppose that $n \geqslant 3$ and $m \geqslant 3$ and that $x_1, ..., x_N$ maximize* $\mathrm{VDM}(x_1, ..., x_N)$ *over $S_{m-1}$. Then* $\max_{x\in S_{m-1}} \sum_{i=1}^{N} l_i^2(x) > 1$.

It is rather peculiar that for $n=2$, examples of such configuration are known for dimensions 2, 6, and 22. A complete list of allowable dimensions would be an interesting curiosity.

Turning our attention to the case of $X=B_m$, Karlin and Studden have shown that there is an optimal measure which is orthogonally invariant and concentrated on a certain number of concentric spheres.

THEOREM 7 [5, p. 354]. *There is an orthogonally invariant probability measure, $\mu^*$, on $B_m$ which maximizes* $\det M(\mu)$. *The radial component is concentrated on a set of $q := \lfloor n/2 \rfloor + 1$ distinct radii $r_1, ..., r_q$. Moreover $r_q = 1$ and $r_1 = 0$ if $n$ is even and $r_1 > 0$ if $n$ is odd.*

We have already considered the case of $n=1$ in general. For $n=2$ consider $m=2$. Then $N=6$, $q=2$ and optimal measure is concentrated at the origin and at the perimeter. This optimal measure is thus of the form given by

$$\int_{B_2} f\,d\mu^* = \alpha f(0,0) + \frac{1-\alpha}{2\pi}\int_0^{2\pi} f(\cos\theta, \sin\theta)\,d\theta.$$

An elementary calculation reveals that $\alpha = \frac{1}{6}$. But

$$\int_0^{2\pi} p(\cos\theta, \sin\theta)\,d\theta = \frac{2\pi}{5}\sum_{k=0}^{4} p(\cos(2k\pi/5), \sin(2k\pi/5))$$

is exact for polynomials of degree 4. Hence one point at the origin together with any five equally spaced on the perimeter will yield $\max_{x\in B_2} \sum_{i=1}^{6} l_i^2(x)$

$=1$ and thus also maximize the Vandermonde determinant. For higher degrees this is in general not possible.

PROPOSITION 8. *Suppose that* $n=3$ *and that* $x_1, ..., x_N \in B_2$ *maximize* $\mathrm{VDM}(x_1, ..., x_N)$. *Then* $\max_{x \in B_2} \sum_{i=1}^{N} l_i^2(x) > 1$.

*Proof.* For $n=3$ and $m=2$, $N=10$. An optimal measure is concentrated on two concentric circles of radii $0 < r_1 < r_2 = 1$ and must be of the form given by

$$\int_{B_2} f\, d\mu^* = \frac{\alpha_1}{2\pi} \int_0^{2\pi} f(r_1 \cos\theta, r_1 \sin\theta)\, d\theta + \frac{\alpha_2}{2\pi} \int_0^{2\pi} f(r_2 \cos\theta, r_2 \sin\theta)\, d\theta.$$

If $\max_{x \in B_2} \sum_{i=1}^{N} l_i^2(x) = 1$, we must have

$$\int_{B_2} p\, d\mu^* = \frac{1}{10} \sum_{i=1}^{10} p(x_i)$$

for all polynomials of degree at most 6. But, if we consider $p = (x^2 + y^2 - r_i^2)\, q$, where $\deg(q) \leqslant 4$, we see that

$$\frac{\alpha_i}{2\pi} \int_0^{2\pi} q(r_i \cos\theta, r_i \sin\theta)\, d\theta = \frac{1}{10} \sum_{|x_k| = r_i} q(x_k), \qquad i = 1, 2.$$

Thus we must have at least five points on each circle. But as there are only ten points all together, we must have exactly five on each circle and they must be equally spaced. Suppose that the points on the $j$th circle are given by

$$(\cos(\varphi_j + 2k\pi/5), \sin(\varphi_j + 2k\pi/5)), \qquad 0 \leqslant k \leqslant 4.$$

Now consider, for fixed $\varphi$, $p = r^5 \sin(5\theta + \varphi)$ given in polar coordinates. Then $\deg(p) = 5$ and we have

$$0 = \frac{1}{10} \left\{ r_1^5 \sum_{k=0}^{4} \sin(5[\varphi_1 + 2k\pi/5] + \varphi) + r_2^5 \sum_{k=0}^{4} \sin(5[\varphi_2 + 2k\pi/5] + \varphi) \right\}$$

$$= \frac{1}{10} \{ 5r_1^5 \sin(5\varphi_1 + \varphi) + 5r_2^5 \sin(5\varphi_2 + \varphi) \}.$$

But this must be true for all $\varphi$, which is not possible as $r_1 < r_2$. ∎

Actually, probably much more is true but we must leave this as a conjecture.

*Conjecture.* Suppose that $x_1, ..., x_N \in B_m$ maximize $\mathrm{VDM}(x_1, ..., x_N)$. Then if $n \geqslant 3$ and $m \geqslant 2$, $\max_{x \in B_2} \sum_{i=1}^{N} l_i^2(x) > 1$.

## References

1. E. Bannai and R. M. Dammerell, Tight spherical designs, I, *J. Math. Soc. Japan* **31** (1979), 199–207.
2. P. Delsarte, J. M. Goethals, and J. J. Seidel, Spherical codes and designs, *Geom. Dedicata* **6** (1977), 363–388.
3. L. Fejér, Bestimmung dergenigen Abszissen eines Intervalles fur welche die Quadratsumme der Grundfunktionen der Lagrangeschen Interpolation im Intervalle ein moglichst kleines Maximum besitzt, *Ann. Scuola Norm. Sup. Pisa* (2) **1** (1932), 263–276.
4. G. H. Golub and C. F. Van Loan, "Matrix Computations," Johns Hopkins Press, Baltimore, 1983.
5. S. Karlin and W. J. Studden, "Tchebycheff Systems: With Applications in Analysis and Statistics," Interscience, New York, 1966.
6. J. Kiefer and J. Wolfowitz, The equivalence of two extremum problems, *Canad. J. Math.* **12** (1960), 363–366.
7. T. Kilgore, A characterization of the Lagrange interpolating projection with minimal Tchebycheff norm, *J. Approx. Theory* **24** (1978), 273–288.