THE APPROXIMATION POWER OF MOVING LEAST-SQUARES

DAVID LEVIN

ABSTRACT. A general method for near-best approximations to functionals on $\mathbb{R}^d$, using scattered-data information is discussed. The method is actually the moving least-squares method, presented by the Backus-Gilbert approach. It is shown that the method works very well for interpolation, smoothing and derivatives' approximations. For the interpolation problem this approach gives Mclain's method. The method is near-best in the sense that the local error is bounded in terms of the error of a local best polynomial approximation. The interpolation approximation in $\mathbb{R}^d$ is shown to be a C^∞ function, and an approximation order result is proven for quasi-uniform sets of data points.

1. INTRODUCTION

Let $f \in F$ where F is a normed function space on $\mathbb{R}^d$, and let $\{L_i(f)\}_{i=1}^I$ be a data set, where $\{L_i\}_{i=1}^I$ are bounded linear functionals on F. In most problems in approximation we are looking for an approximation to $L(f)$, where L is another bounded linear functional on F, in terms of the given data $\{L_i(f)\}_{i=1}^I$. Usually we choose a set of basis functions, $\{\phi_k\} \subset F$, e.g., polynomials, splines, or radial basis functions. Then we find an approximation $\widehat{f}$ to f from span$\{\phi_k\}$, and approximate $L(f)$ by $L(\widehat{f})$. If the approximation process is linear, the final approximation can be expressed as

$$(1.1) \qquad \widehat{L}(f) \equiv L(\widehat{f}) = \sum_{i=1}^I a_i L_i(f).$$

In analyzing the approximation error, or the approximation order, we are frequently using the fact that the approximation procedure is exact for a finite set of fundamental functions $P \equiv \text{span}\{p_j\}_{j=1}^J \subset F$ (usually polynomials)

$$(1.2) \qquad \widehat{L}(p) = \sum_{i=1}^I a_i L_i(p) = p, \quad p \in P.$$

In case the basis functions $\{\phi_k\} \subset F$ are locally supported and $P = \Pi_m$, it can be shown, in many problems, that the resulting approximation is $O(h^{m+1})$, where h is a local data parameter. Another way of analyzing the approximation error follows directly from the representation (1.1).

Let $\Omega_0 \subset \mathbb{R}^d$ be the support of the functional L, i.e., $L(g) = 0$ for all g vanishing on Ω_0, and let Ω_I denote the support of $\sum_{i=1}^I a_i L_i$. Also let p be the best

approximation to f from the set P on $\Omega \equiv \Omega_0 \cup \Omega_I$,

$$(1.3) \qquad\qquad E_{\Omega,P}(f) \equiv \|f - p\|_\Omega = \inf_{q \in P} \|f - q\|_\Omega,$$

where $\|\cdot\|_\Omega$ is the natural restriction to Ω of the norm on F. Using (1.2) it follows that

$$|L(f) - \widehat{L}(f)| \le |L(f) - L(p)| + |L(p) - \widehat{L}(f)|$$

$$\le \|L\|\|f - p\|_\Omega + |\sum_{i=1}^{I} a_i L_i(f - p)|$$

$$(1.4)$$

$$\le \|L\|\|f - p\|_\Omega + \sum_{i=1}^{I} |a_i|\|L_i\|\|f - p\|_\Omega$$

$$= (\|L\| + \sum_{i=1}^{I} |a_i|\|L_i\|)E_{\Omega,P}(f).$$

Thus, a bound on the error in the approximation (1.1) is given in terms of the norms of L and the $L_i's$, the coefficients $\{a_i\}_{i=1}^{I}$, and of $E_{\Omega,P}(f)$, the error of best approximation to f on Ω from P. Similarly, let $\widetilde{p}$ be the best approximation to f on Ω from all $q \in P$ such that $L(q) = L(f)$, and let

$$(1.5) \qquad\qquad E_{\Omega,P}^{L}(f) \equiv \|f - \widetilde{p}\|_\Omega = \inf_{q \in P, L(q)=L(f)} \|f - q\|_\Omega.$$

Then it follows that

$$(1.6) \qquad\qquad |L(f) - \widehat{L}(f)| \le \sum_{i=1}^{I} |a_i|\|L_i\|E_{\Omega,P}^{L}(f).$$

In this work we skip the phase of approximating f by $\widehat{f} \in \text{span}\{\phi_k\}$. We directly construct approximations of the form (1.1) which satisfy (1.2). It is clear from (1.4) that on one hand we would like to minimize $\sum_{i=1}^{I} |a_i|\|L_i\|$, and on the other hand we want a minimal domain $\Omega = \Omega_0 \cup \Omega_I$, where Ω_I is the support of $\sum_{i=1}^{I} a_i L_i$. The second property is for reducing the factor $E_{\Omega,P}(f)$ in (1.4). If we could achieve a good solution satisfying those two interrelated goals, then we would get an approximation which is "near-best", in the sense that its error is bounded by a small factor times the error of the best local approximation from P. In Section 2 we present a constructive way of computing such approximations. The general idea is motivated here through approximation theory considerations. However, it brings us back to the Backus-Gilbert theory [BG1]-[BG3], [BS]. The resulting approximants are closely related to the moving least-squares methods, starting with Shepard's method [Sh] and its extensions by McLain [Mc1], [Mc2], Franke and Nielson [FrNi], and Lancaster and Salkauskas [LS]. The generalized moving least-squares method is defined as follows.

Definition. The moving least-squares approximation. Let $\{L_i\}_{i=1}^{I}$, L and P be as above, and let $\{\Theta(L_i, L)\}$ be some non-negative weights. The moving least-squares approximation to $L(f)$ is defined as $L(p^*)$, where $p^* \in P$ is minimizing, among all $p \in P$, the weighted least-squares error functional

$$(1.7) \qquad\qquad \sum_{i=1}^{I}(L_i(p) - L_i(f))^2\Theta(L_i, L).$$

The moving least-squares method is better known for function value approximation from pointwise scattered data:

Let $\{x_i\}_{i=1}^I$ be a set of distinct data points in $\mathbb{R}^d$, and let $\{f(x_i)\}_{i=1}^I$ be the data values at these points. The moving least-squares approximation of order m at a point $x \in \mathbb{R}^d$ is the value $p^*(x)$ where $p^* \in \Pi_m$ is minimizing, among all $p \in \Pi_m$, the weighted least-squares error

$$(1.8) \qquad \sum_{i=1}^I (p(x_i) - f(x_i))^2 \theta(\|x - x_i\|),$$

where θ is a non-negative weight function and $\|\cdot\|$ is the Euclidean distance in $\mathbb{R}^d$. The approximation is made local if $\theta(r)$ is fast decreasing as $r \to \infty$ and interpolation is achieved if $\theta(0) = \infty$.

The connection between the moving least-squares method and the Backus-Gilbert theory is shown by Abramovici [Ab] for Shepard's method, and for the general case by Bos and Salkauskas [BS]. This connection is re-established here using a simpler approach. We then show that the suggested approximants are C^∞, and prove a near-best property and an approximation order result. In Section 3 we apply the method to univariate interpolation, smoothing and derivative approximation, and to scattered-data interpolation and to derivative approximation in $\mathbb{R}^d$, $d = 2, 3$. Further, we present some ideas for data dependent procedures for multivariate scattered-data interpolation.

2. THE APPROXIMATION SCHEME
AND APPLICATION TO SCATTERED-DATA INTERPOLATION

Let us assume that the data set $\{L_i(f)\}_{i=1}^I$ is finite and $J \leq I$. We suggest finding the coefficients vector $\overline{a} = (a_1, \ldots, a_I)^t$ for the approximation $\widehat{L}(f) = \sum_{i=1}^I a_i L_i(f)$ by minimizing the quadratic form

$$(2.1) \qquad Q = \sum_{i=1}^I w(L_i, L) a_i^2,$$

subject to the linear constraints

$$(2.2) \qquad \sum_{i=1}^I a_i L_i(p_j) = L(p_j), \quad j = 1, \ldots, J.$$

In (2.1) $\{w(L_i, L)\}$ are non-negative weights, $w(L_i, L)$ represents a separation measure between the functionals L_i and L, $w(L_i, L) > 0$ if $L_i \neq L$.

Proposition 1. 1. *Assume $L \neq L_k$, $k = 1, \ldots, I$, and $\mathrm{Rank}(E) = J$ where $E_{i,j} = L_i(p_j)$ for $1 \leq i \leq I$, $1 \leq j \leq J$. Then the approximation defined by the constraint minimization problem (2.1)-(2.2) is $\widehat{L}(f) = \sum_{i=1}^I a_i L_i(f) = \overline{\sigma}^t \overline{a}$ with $\overline{\sigma} = (L_1(f), \ldots, L_J(f))^t$ and*

$$(2.3) \qquad \overline{a} = D^{-1} E (E^t D^{-1} E)^{-1} \overline{c},$$

where $D = 2 \, \mathrm{Diag}\{w(L_1, L), \ldots, w(L_I, L)\}$ and $\overline{c} = (L(p_1), \ldots, L(p_J))^t$.
2. *If the separation measure w satisfies $w(L, L) = 0$ the suggested scheme is interpolatory in the sense that if $L = L_k$ for some $1 \leq k \leq I$, then $\widehat{L}(f) = L(f)$.*

3. *The moving least-squares approximation defined by minimizing* (1.7), *is the same as the approximation obtained by the constrained least-squares problem* (2.1)-(2.2) *with* $w(L_i, L) = \Theta(L_i, L)^{-1}$.

Proof. The constraint minimization problem (2.1)-(2.2) is transformed, using Lagrange multipliers $z_1, \ldots, z_J$, into the linear system

$$(2.4) \qquad D\overline{a} + E\overline{z} = 0, \quad \text{and} \quad E^t\overline{a} = \overline{c}.$$

It can easily be verified that the matrix of the system,

$$(2.5) \qquad A = \begin{pmatrix} D & E \\ E^t & 0 \end{pmatrix},$$

is non-singular, and that $\overline{a} = D^{-1}E(E^tD^{-1}E)^{-1}\overline{c}$ and $\overline{z} = (E^tD^{-1}E)^{-1}\overline{c}$ solve (2.4). The second claim follows by observing that $a_i = \delta_{i,k}$ solves (2.2) and gives $Q = 0$.

Choosing $\Theta(L_i, L) = w(L_i, L)^{-1}$ in (1.7) and expressing the functions in P by the basis $\{p_j\}_{j=1}^J$, it directly follows that the minimizer of (1.7) is $p^* = \sum_{j=1}^J b_j p_j$ with

$$(b_1, \ldots, b_J) = \overline{\sigma}^t D^{-1}E(E^tD^{-1}E)^{-1}.$$

Thus $L(p^*) = (b_1, \ldots, b_J)\overline{c} = \overline{\sigma}^t\overline{a} = \widehat{L}(f)$ and the third assertion is proved. $\qquad \square$

The Backus-Gilbert approach [BG1]-[BG3] is exactly in the form of the constrained least-squares problem (2.1)-(2.2), with only $J = 1$ in (2.2) and $p_1 \equiv 1$. The general case is discussed in [BS]. Part 3 of Proposition 1 reasserts the connection between the moving least-squares method and the Backus-Gilbert optimality which was shown by Abramovici [Ab] for Shepard's method, and for the more general case by Bos and Salkauskas [BS]. The proof here is much simpler.

Let us demonstrate the application of the above general scheme to the problem of scattered-data interpolation in $\mathbb{R}^d$. Given a set of data values $\{f(x_i)\}_{i=1}^I$ one would like to find an approximating function $\widehat{f}(x)$ satisfying $\widehat{f}(x_i) = f(x_i)$, $i = 1, \ldots, I$, and such that $\widehat{f}$ is a good approximation to the underlying function f. Many methods were suggested for solving this problem, see e.g. [Fr], all based on explicit functional forms for $\widehat{f}$. In our terminology, we are given the values of the evaluation functionals $L_i(f) = f(x_i)$, $i = 1, \ldots, I$, and we look for an approximation to the evaluation functional $L(f) = f(x)$ at a given $x \in \mathbb{R}^d$. For evaluation functionals we take radial separation measures $\{w(L_i, L)\}$ of the form

$$(2.6) \qquad w(L_i, L) = \eta(\|x - x_i\|),$$

where η is an increasing function on $\mathbb{R}_+$, and $\eta(0) = 0$. The support of the functional $\sum_{i=1}^I a_i L_i$ here is simply set $\Omega_I = \{x_i \mid a_i \neq 0\}$, or practically, $\Omega_I = \{x_i \mid |a_i| > \varepsilon\}$, where ε is a small positive constant. Here $\Omega_0 = \{x\}$, so to minimize $\Omega = \Omega_0 \cup \Omega_I$ we would like to make Ω_I as close as possible to x. Hence we assign very large penalty weights to points which are far from x. Having tried several alternatives for the function η, we found out that a very good choice is

$$(2.7) \qquad \eta(r) = \exp(\frac{r^2}{h^2}) - 1,$$

where h is an average distance between the data points. In Section 4 we consider non-radial data-dependent separation measures.

The interpolation procedure in $\mathbb{R}^d$. Given data values $\{f(x_i)\}_{i=1}^I$, where $\{x_i\} \subset \mathbb{R}^d$ is a set of distinct data points, we define the approximation at $x \in \mathbb{R}^d$ as

$$(2.8) \qquad \widehat{L}(f) = \sum_{i=1}^I a_i f(x_i),$$

where $\{a_i\}$ minimizes

$$(2.9) \qquad Q = \sum_{i=1}^I \eta(\|x - x_i\|) a_i^2,$$

subject to the constraints

$$(2.10) \qquad \sum_{i=1}^I a_i p_j(x_i) = p_j(x), \quad j = 1, \dots, J,$$

where $\{p_j\}_{j=1}^J$ span Π_m, the space of polynomials of total degree $\leq m$, $J = \binom{d+m}{m}$. Here the system (2.4) is defined with with $D = 2\,\mathrm{Diag}\{\eta(\|x-x_1\|), \dots, \eta(\|x-x_I\|)\}$, $E_{i,j} = p_j(x_i)$, $1 \leq i \leq I, 1 \leq j \leq J$ and $\overline{c} = (p_1(x), \dots, p_J(x))^t$. For $J = 1$ and $\eta(r) = r^2$, it follows ([Ab]) that the resulting approximation is the well known Shepard interpolant. For $J > 1$ and $\eta(r) = r^2 \exp(\alpha r^2)$, we rediscover McLain's method [Mc1], [Mc2].

Proposition 2. *The matrix A, given by* (2.5) *with D and E defined above and η given by* (2.7)*, is non-singular if* $\mathrm{Rank}(E) = J$ *and* $1 \in P$.

Proof. Assume $\overline{y} = (\overline{a}, \overline{z})$ is a solution of $A\overline{y} = 0$, i.e.,

$$(2.11) \qquad D\overline{a} + E\overline{z} = 0, \quad \text{and} \quad E^t \overline{a} = 0.$$

Multiplying (2.11) by $\overline{a}^t$ we obtain $0 = \overline{a}^t D\overline{a} + \overline{a}^t E\overline{z} = \overline{a}^t D\overline{a} + \overline{z}^t E^t \overline{a} = \overline{a}^t D\overline{a}$. If $x \neq x_i$, $1 \leq i \leq I$, then $\overline{a}^t D\overline{a} = 0$ implies $\overline{a} = 0$, and since $\mathrm{Rank}(E) = J$ it follows also that $\overline{z} = 0$. If $x = x_i$ for some $1 \leq i \leq I$, then the ith component of $\overline{a}$ may be non-zero. However, since $1 \in P$, this contradicts $E^t \overline{a} = 0$. $\qquad\square$

The above results lead to the following interpolation theorem (see also [LS]).

Theorem 3. *Let $P \subset C^\infty(\mathbb{R}^d)$, $1 \in P$, and let the distribution of data points $\{x_i\}$ be such that $\mathrm{Rank}(E) = J$. Then the approximation $\sum_{i=1}^I a_i f(x_i)$ is a C^∞ function of x, interpolating the data $\{f(x_i)\}$. Moreover, the coefficients $\{a_i\}$ decay exponentially in $\|x - x_i\|^2 / h^2$, i.e.*

$$(2.12) \qquad |a_i| \leq K(\exp(-\frac{\|x - x_i\|^2}{h^2}))(\|x - x_i\|/h)^m,$$

for $\|x - x_i\|/h > 1$, where $K = J\|(E^t D^{-1} E)^{-1}\|$.

Proof. The first part follows by observing that the elements of A^{-1} and $\overline{c}$ are in $C^\infty(\mathbb{R}^d)$. Hence the solution vector $\overline{a}$ of the system (2.4) is analytic in x. The decay rate is derived by the explicit form (2.3) for $\overline{a}$. $\qquad\square$

Computation of the interpolant. Practical considerations. The interpolant of order m is defined by (2.8) where the coefficient vector $\overline{a}$ is defined by (2.3) with η given by (2.7), $P = \Pi_m$, $D = 2\,\text{Diag}\{\eta(\|x - x_1\|), \dots, \eta(\|x - x_I\|)\}$, $E_{i,j} = p_j(x_i)$, $1 \le i \le I, 1 \le j \le J$ and $\overline{c} = (p_1(x), \dots, p_J(x))^t$.

Since $P = \Pi_m$ is shift invariant, we suggest taking the basic functions $\{p_j\}$ to be the monomials shifted to x. Thus we avoid using large numbers in E and we have $\overline{c} = (1, 0, \dots, 0)$. Note that the computation of the moving least-squares interpolant of order m involves the solution of a Jth order system of equations, $J = \dim \Pi_m$, and in most applications $J << I$.

To avoid working with very large numbers in D, as the elements $\{\eta(\|x - x_i\|)\}$ can be, we normalize the system of equations as follows. We divide row i and column i of A by $\sqrt{1 + D_{ii}}$ and obtain the system $\widetilde{A}\widetilde{a} = \overline{b}$, where $\widetilde{a}_i = a_i \cdot \sqrt{1 + D_{ii}}$ and

$$(2.13) \qquad \widetilde{A} = \left(\begin{smallmatrix} \widetilde{D} & \widetilde{E} \\ \widetilde{E}^t & 0 \end{smallmatrix} \right),$$

with $\widetilde{D}_{ii} = D_{ii}/(1 + D_{ii})$ and $\widetilde{E}_{i,j} = E_{i,j}/\sqrt{1 + D_{ii}}$, $1 \le i \le I, 1 \le j \le J$.

Localized moving least-squares interpolants. The weight function η defined by (2.7) highly penalizes the use of points far from x. However, to obtain an approximation functional of a finite support one should use a function η which is infinite outside a certain circle around x, say of radius s. Also, we like to maintain the nice property of the approximant stated in Theorem 3, i.e., to make sure that the resulting approximation is C^∞. Both these objectives can be achieved with a properly chosen η, as stated in the following:

Theorem 4. *Let $S \subset \mathbb{R}^d$ and $s > 0$ be such that $\text{Rank}(\widetilde{E}) = J$ for any $x \in S$, where*

$$(2.14) \qquad \widetilde{E}_{i,j} = p_j(x_i) \cdot \chi_{x,s}(x_i),$$

for $1 \le i \le I, 1 \le j \le J$, and $\chi_{x,s}$ is the characteristic function of the open ball of radius s around x, $B(x, s)$. Also let

$$(2.15) \qquad \eta^{-1}(r) = (\exp(\frac{r^2}{h^2}) - 1)^{-1} \cdot \exp(-\frac{s^2}{(s-r)^2}) \cdot \chi_{0,s}(r).$$

Then,

1. *The approximation defined by (2.8)-(2.10) is a C^∞ function of x on S, interpolating the data $\{f(x_i)\}$.*
2. *The approximation is local, i.e., at each point $x \in S$ it uses only data values within the open ball of radius s around x.*
3. *The error in the approximation can be bounded in terms of the coefficients $\{a_i\}$ and the error of the best local approximation to f from Π_m,*

$$(2.16) \qquad |f(x) - \sum_{i=1}^{I} a_i f(x_i)| \le (1 + \sum_{i=1}^{I} |a_i|) E_{B(x,s),\Pi_m}(f).$$

Proof. The above remark on normalizing the system of equations is useful here. The matrices $\widetilde{D}$ and $\widetilde{E}$ are modified by assigning $\widetilde{D}_{ii} = 1$ and $\widetilde{E}_{i,j} = 0$, $1 \le j \le J$, if $\chi_{x,s}(x_i) = 0$. The matrix $\widetilde{A}$ is nonsingular for any $x \in S$, and the solution vector $\overline{a}$ is C^∞ as a function of x on S. It is also clear that $\widetilde{a}_i = 0$, and thus also $a_i = 0$, if $\chi_{x,s}(x_i) = 0$. The error bound (2.16) is derived from (1.4). $\qquad \square$

The expression (2.16) for the error indicates the near-best property of the suggested approximant. We know that $\sum_{i=1}^{I} |a_i| \geq 1$ since $\sum_{i=1}^{I} a_i = 1$. Therefore, if in application we get that $\sum_{i=1}^{I} |a_i|$ is not much greater than 1, then we know that the error in the approximation is near the error in the local best approximation to f from Π_m. We demonstrate this in Section 3.

The interpolant proposed in Theorem 4 is of local support, and it is exact for $f \in \Pi_m$. Therefore, it is expected to give an $O(h^{m+1})$ approximation order, provided that f is sufficiently smooth. A first result of this type is presented by Farwig in [Fa1] for approximations obtained by using Shepard's interpolation to local truncated Taylor series data. Further important results on the analysis of moving least-squares methods are presented by Farwig in [Fa2]. The above result (2.16) can also be derived by the analysis in [Fa2]. For presenting an approximation order analysis we first need, as in [BDL], a proper definition of a "mesh size" for scattered data points in $\Omega \subset \mathbb{R}^d$.

Definition. h-ρ-δ sets of mesh size h, density $\leq \rho$, and separation $\geq \delta$.
Let Ω be a domain in $\mathbb{R}^d$, and consider sets of data points in Ω. We say that the set $X = \{x_i\}_{i=1}^{I}$ is an h-ρ-δ set if

1. h is the minimal number such that

$$(2.17) \qquad \Omega \subset \cup_{i=1}^{I} \overline{B}(x_i, h/2),$$

   where $\overline{B}(x, r)$ is the closed ball of radius r with center x.

2.

$$(2.18) \qquad \#\{X \cap \overline{B}(y, qh)\} \leq \rho \cdot q^d, \quad q \geq 1, \quad y \in \mathbb{R}^d.$$

   Here, $\#\{Y\}$ denotes the number of elements of a given set Y.

3. $\exists \, \delta > 0$ such that

$$(2.19) \qquad \|x_i - x_j\| \geq h\delta, \quad 1 \leq i < j \leq I.$$

It is quite easy to prove an $O(h^{m+1})$ approximation order for h-ρ-δ data sets if the definition of the approximant is revised as follows. Instead of minimizing (2.9) we minimize

$$(2.20) \qquad Q = \sum_{i=1}^{I} |a_i|,$$

with

$$(2.21) \qquad a_i = 0, \quad \text{if } x_i \notin \overline{B}(x, qh),$$

where q is a fixed positive number. This definition is also more appropriate in view of the expression (2.16) for the error (with $S = \overline{B}(x, qh)$). However, the resulting approximant would not even be continuous in x. In Theorem 5 we prove the required $O(h^{m+1})$ result for the smooth approximant defined by (2.8)-(2.10), and for h-ρ-δ data sets.

Theorem 5. *Let f be a function in $C^{m+1}(\Omega)$. Then, for fixed ρ and δ, there exists a fixed $q > 0$, independent of h, such that the approximant $\widehat{L}(f)$ defined by (2.8)-(2.10), with η given by (2.15) and $s = qh$, satisfies*

$$(2.22) \qquad \|\widehat{L}(f) - f\|_{\Omega, \infty} < M \cdot h^{m+1},$$

for h sufficiently small, for h-ρ-δ sets of data points.

Proof. As in [BDL] it can be shown that, for sufficiently small h, there exists a fixed $q > 0$ such that the conditions of Theorem 4 hold with $s = qh$ and $S = \Omega$ for all sets of data points of mesh size h. For sets of density $\leq \rho$ the number of data points in $\overline{B}(x, qh)$ is bounded by $N = \rho \cdot q^d$, independent of h. So there are at most N coefficients $\{a_i\}$ participating in the approximant. Now we observe that the coefficients $\{a_i\}$ are continuous functions of the data points $\{x_i\}$. Thus, for a fixed x and a fixed h it follows that $\sum\limits_{x_i \in \overline{B}(x,qh)} |a_i| \leq C$, where C is the same constant for all h-ρ-δ sets of data points. Also it is clear that C does not depend on x. We further observe that the approximant in hand is invariant under the contraction mapping $x \rightarrow \alpha x$, $x_i \rightarrow \alpha x_i$ and $h \rightarrow \alpha h$. Then it follows that C is also independent of h. The proof is completed by estimating the error as in (1.4) with p being the Taylor expansion of total degree m of f at x. $\qquad\square$

We believe that the above approximation results can be extended as follows.

1. The restrictions on the density and the separation in Theorem 5 may be removed.
2. For h-ρ-δ sets of data points the constant K in (2.12) is uniformly bounded in x.
3. The approximation rate result (2.22) holds with η given by (2.7).

3. Numerical demonstration

The moving least-squares method can be applied to many problems in multivariate approximation, and it calls for a comprehensive testing and comparison study. In this section we present some results on the application of the method to interpolation, smoothing and derivative approximation in $\mathbb{R}^1$, and to scattered-data interpolation in $\mathbb{R}^d$, $d = 2, 3$. Unlike the experiments in [Fr], here we mainly tested the locality and the norm of the approximation coefficients' vector $\overline{a}$. We hope to convince the reader that the method gives "near-best" approximations. Moreover, from the experience gained with this approximation, we feel that this rather simple method gives almost the best result among all linear methods based on the given data.

Example 1. Univariate interpolation and smoothing. Consider the very simple case of univariate interpolation at equidistant points in $[0, 1]$, $x_i = 0.1(i-1)$, $i = 1, \ldots, 11$. Consider the moving least-squares interpolant defined by (2.7) with $h = 0.1$. The coefficients $\{a_i\}$ were computed by (2.3), which involve the solution of a linear system of order $J = m + 1$. For $m = 2$ the approximant reproduces quadratic polynomials. To demonstrate its locality and its "near-best" property, let us view the behavior of the coefficients $\{a_i\}$ and their ℓ_1-norm $\|\overline{a}\|_1 = \sum\limits_{i=1}^{I} |a_i|$ as functions of x in $[0, 1]$. We find that $\max_{x \in [0,1]} \|\overline{a}\|_1 < 1.24$ which implies, by the remark following Theorem 4, that the error is bounded by 2.24 times the error of the best approximation by a quadratic polynomial. In Figure 1 we give graphs of $\|\overline{a}\|_1$ as a function of x, $x \in [0, 1]$. The upper graph is for the case of 11 equidistant points in $[0, 1]$, and the lower graph is for 11 randomly chosen points in $[0, 1]$. The points with $\|\overline{a}\|_1 = 1$ are, of course, the interpolation points. The approximation is also found to be very local. For example, the coefficients' vector $\overline{a}$ for approximation at the point $x = 0.33$, by the above 11 equidistant points, is

$\overline{a} = (-4.22 \cdot 10^{-5}, -5.69 \cdot 10^{-3}, -7.73 \cdot 10^{-2}, 8.62 \cdot 10^{-1}, 2.30 \cdot 10^{-1}, 8.73 \cdot 10^{-3}, -5.47 \cdot 10^{-4}, -2.05 \cdot 10^{-6}, -8.11 \cdot 10^{-10}, -3.81 \cdot 10^{-14}, -2.23 \cdot 10^{-19})^t$.

We also applied the interpolation procedure to a specific function. To demonstrate the power of the method and the significance of the strong locality, we choose the function

$$(3.1) \qquad f(x) = (2 - \cos(10x))|1 - 4x^2|H(x - 0.75),$$

shown in Figure 2. This function has 5 local extrema, a point of jump discontinuity, a point of derivative discontinuity, and a flat part. This "rich" function is measured at 11 equidistant points in $[0, 1]$ and the interpolant is computed with $m = 2$. The resulting approximant is depicted in the upper graph in Figure 3. It evidently shows the very localized influence of the points of discontinuity, and the power of reconstructing the major features of the approximated function. In the lower graph in Figure 3 we demonstrate the potential application of the moving least-squares method for smoothing. The smoothing here is obtained by using the method with η defined as

$$(3.2) \qquad \eta(r) = \exp(\frac{r^2}{h^2}).$$

Example 2. Univariate derivative approximation. Next we consider approximation to the derivative of functions in $[0, 1]$, based on function values at the same 11 points as above, $x_i = 0.1(i - 1)$, $i = 1, \ldots, 11$. Here we are given the functionals $L_i(f) = f(x_i)$ and we look for an approximation to $L(f) = f'(x)$. Since we are not looking for interpolation, we define the separation measure here by (2.6) with η defined by (3.2). We took $P = \Pi_4$ and computed the coefficients of the approximant, for 101 equidistant points in $[0, 1]$, by equation (2.3). We find that

$$\max_{x \in [0.2, 0.8]} \|\overline{a}\|_1 < 22,$$

and $\max_{x \in [0,1]} \|\overline{a}\|_1 < 107$. These bounds are not very big since for derivative schemes one expects $\|\overline{a}\|_1 = O(h^{-1})$. Figure 4 shows the graph of $\|\overline{a}\|_1$ as a function of x.

The error in the approximation to the derivative of $f(x) = \cos(x)$ at 101 equidistant points in $[0, 1]$ is shown in Figure 5. The main features observed here are that the absolute value of the error gets bigger toward the ends of the interval, as $\|\overline{a}\|_1$ does, and it has local maxima near data points and local minima near midpoints.

Example 3. Scattered data interpolation in $\mathbb{R}^2$ and $\mathbb{R}^3$. The common way of presenting the efficiency of a multivariate approximation method is by applying it to many test functions and listing the resulting errors. Following the general idea of this work, driven by the error bound (1.4), it suffices to present bounds on $\|\overline{a}\|_1$, and to demonstrate the locality property of the coefficients $\{a_i\}$.

We examined interpolation at data points scattered in $[0, 1]^d$, for $d = 2, 3$, and obtained very good approximations for many test functions. We choose to demonstrate here that we really achieve the goals set in the Introduction. Namely, a smooth approximation scheme of the form (1.1) where the norm $\|\overline{a}\|_1$ of the coefficients is small, and also the support of the coefficients is small. We applied the moving least-squares interpolant of order 3, for a set of 81 data points randomly chosen in $[0, 1]^2$, with $h = 1/9$. In Figure 6 we display the coefficients of the approximant at the point $x = (0.75, 0.75)$ (depicted by a cross), where each point is

represented by a circle of diameter proportional to the absolute value of the corresponding coefficient (plus a very small constant). The circles are open to the right if the coefficients are positive, and to the left if negative. We observe that the significant coefficients are all positive, and that they are all in a small neighborhood of the point $x = (0.75, 0.75)$. It turns out that all the coefficients of absolute value > 0.01 are within a circle of radius 0.25 around $(0.75, 0.75)$ and that $\|\overline{a}\|_1 = 1.462$. Thus we know, by (1.4), that the error is bounded by 2.462 times the error of the best approximation to f from Π_3 in a small neighborhood of x. For comparison we evaluated the polynomial approximation at $x = (0.75, 0.75)$ based on interpolation at the 10 closest points to x from the same set of data points $(\dim \Pi_3 = 10)$. If we express this approximation in the form (2.8), it turns out that $\|\overline{a}\|_1 = 204.24$. thus we conclude that this local interpolation approach, apart from being non-smooth, yields much worse approximations.

Similar properties were observed in the 3-D case. Here we applied the moving least-squares interpolant of order 2, for a set of 125 data points randomly chosen in $[0, 1]^3$, with $h = 0.2$. In Figure 7 we display the coefficients of the approximant at the point $x = (0.25, 0.25, 0.25)$ (depicted by a 3-D cross), where each point is represented by a box of a size proportional to the value of the corresponding coefficient. The boxes are open above if the coefficients are positive, and open below if negative. We observe that the significant coefficients are all positive and that they are all in the small neighborhood of the point $x = (0.25, 0.25, 0.25)$. For this point we have $\|\overline{a}\|_1 = 1.345$, so again the approximation is local near-best. For example, for the specific function $f(x) = \cos(2x)e^y(x + y + z)^2$, the error at that point is ≈ 0.001.

We remark that the local scheme suggested in Theorem 4 has been tested and the results do not differ much from the global method.

4. Toward data dependent approximants

The procedures discussed above are linear, and they do not use any information about the approximated function. We know that data dependent approximations can be much better, see [DLR] for example. With the approximation scheme presented here we have a convenient way of introducing a good data dependent scheme. The essential point in data dependent approximation in $\mathbb{R}^d$ is to use more information from directions in which the variations in the function are small, and less information from directions of high variations. The variations can be measured by the second order partial derivatives of the function. We can adapt our definition of the interpolation procedure (2.7)-(2.10) by considering a non-radial penalty function,

$$(4.1) \qquad\qquad w(L_i, L) = \exp((x - x_i)^t T(x - x_i)/h^2) - 1,$$

to be used in (2.1). Here T is a $d \times d$ positive definite matrix, to be chosen so that the ellipsoid $y^t T y = 1$ has larger axes in the directions of smaller variations in the function. A method for determining these directions may be based upon estimating the second order partial derivatives of the function by the linear method presented in this paper. We have performed some numerical experiments based on this idea, and it proved to be very good. Here we choose to present just the influence of introducing the matrix T on the coefficients of the approximant. We return to the above two dimensional example, which uses $T = I$, and, for the same data points

and for $x = (0.75, 0.75)$, we replaced the penalty function by (4.1) with

$$(4.2) \qquad\qquad T = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0.05 \end{smallmatrix} \right).$$

The resulting coefficients are displayed in Figure 8, using the same display code as in Figure 6. It is clearly seen that the approximation now uses more information from the y direction, and that the norm of the coefficients is still small, $\|\overline{a}\|_1 = 1.34$.

If we want to use the above idea of directional interpolation and still maintain the high smoothness of the approximant, the matrix T should be smooth in x. Another idea, which has not been tested, is to use a variable parameter h in the definition of the approximant. This can be useful for cases where the density of data points is highly non-uniform, with h representing the local mesh size. As in the previous case, h should vary smoothly in x. Another issue that has not been tested enough is the use of sets P other than polynomials, e.g., trigonometric functions, or splines.
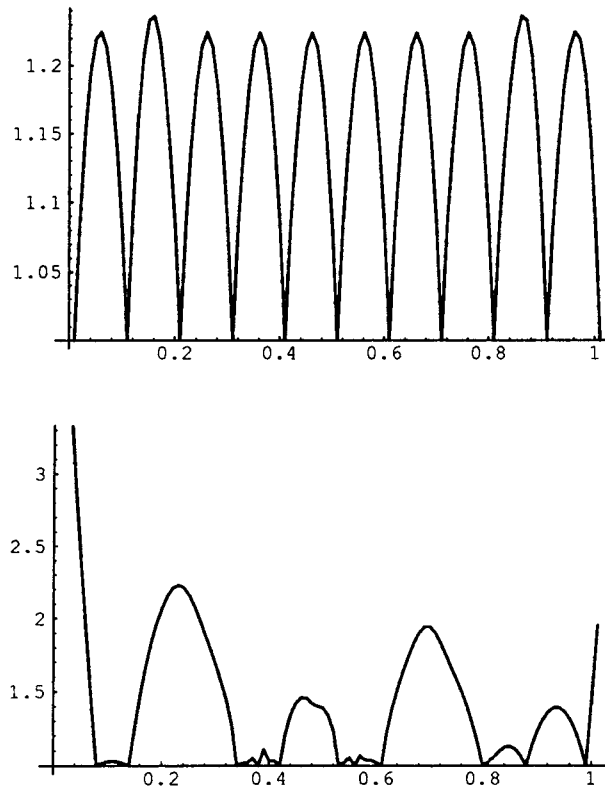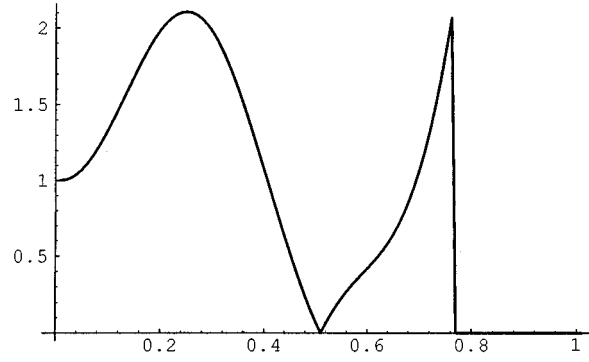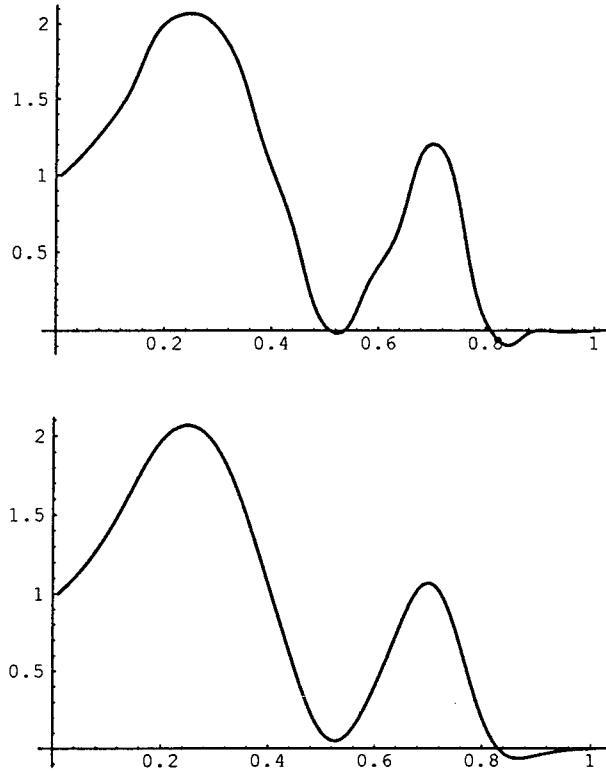

FIGURE 1. The graph of $\|\overline{a}\|_1$ as a function of x for 11-point interpolation in [0,1], with equidistant points (up) and scattered points (down).

FIGURE 2. The test function f

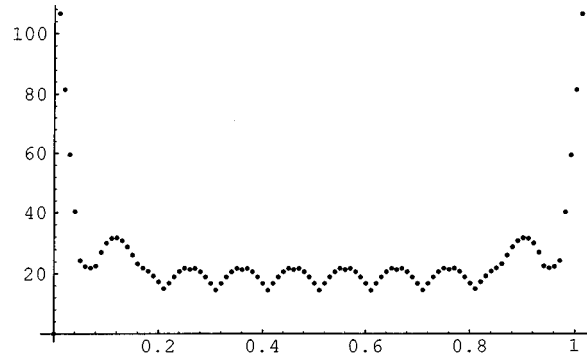FIGURE 3. The 11-point moving least-squares interpolation to f (up) and smoothing (down).

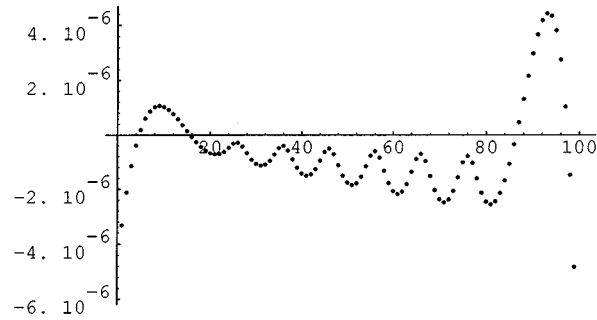FIGURE 4. The graph of $\|\bar{a}\|_1$ (for derivative approximation) as a function of x.

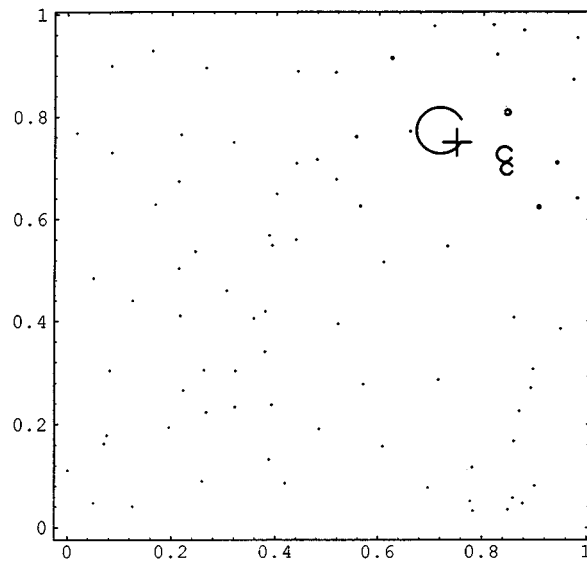FIGURE 5. The error in the approximation to $f'(x)$, $f(x) = \cos(x)$.

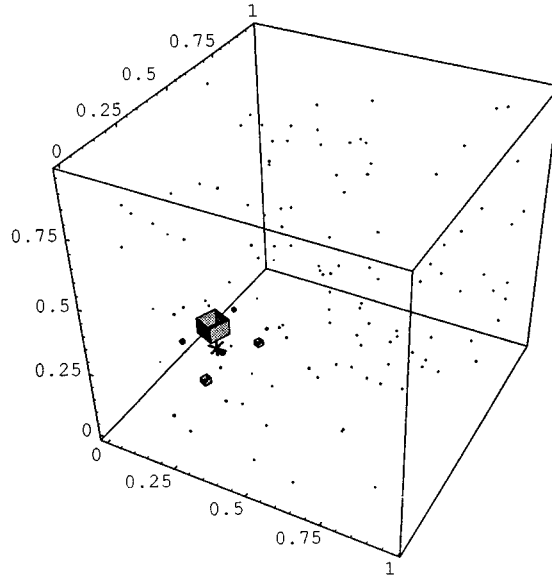FIGURE 6. The coefficients $\{a_i\}$ for interpolation at $x = (.75, .75)$.

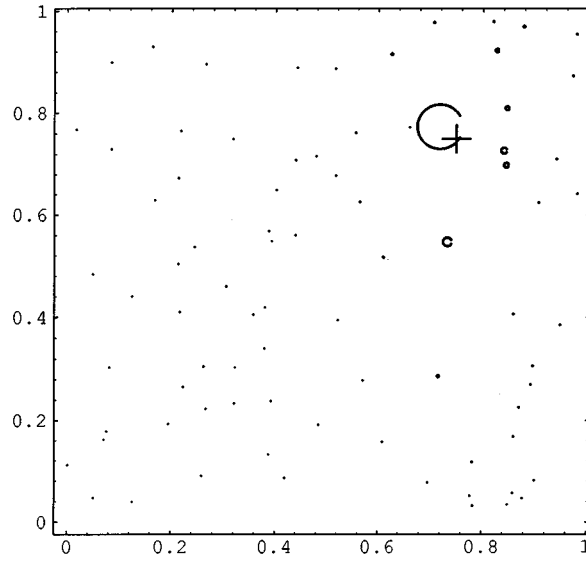FIGURE 7. The coefficients $\{a_i\}$ for interpolation at $x = (.25, .25, .25)$.

FIGURE 8. The coefficients for directional interpolation at $x = (0.75, 0.75)$.

References

[Ab]   F. Abramovici, 1984 *The Shepard interpolation as the best average of a set of data*, Technical Report, Tel-Aviv University.

[BDL]   M.D. Buhmann, N. Dyn and D. Levin, 1995 *On quasi-interpolation by radial basis functions with scattered centers*, Constructive Approximation **11** 239-254. MR **96h:**41038

[BG1]   G. Backus and F. Gilbert, 1967 *Numerical applications of a formalism for geophysical inverse problems*, Geophys. J.R. Astr. Soc. **13** 247-276.

[BG2]   G. Backus and F. Gilbert, 1968 *The resolving power of gross Earth data*, Geophys. J.R. Astr. Soc. **16** 169-205.

[BG3]   G. Backus and F. Gilbert, 1970 *Uniqueness in the inversion of inaccurate gross Earth data*, Philos. Trans. Roy. Soc. London, Ser. A, **266** 123-192. MR **56:**7763

[BS]   L. Bos and K. Salkauskas, 1989 *Moving least-squares are Backus-Gilbert optimal*, J. Approx. Theory **59** 267-275. MR **91a:**41003

[DLR]   N. Dyn, D. Levin and S. Rippa, 1990 *Data dependent triangulation for piecewise linear interpolation*, IMA J. Numer. Anal. **10** 137-154. MR **91a:**65022

[Fa1]   R. Farwig, 1986 *Rate of convergence of Shepard's global interpolation formula*, Math. Comp. **46**, No. 174, 577-590. MR **88a:**65015

[Fa2]   R. Farwig, 1986 *Multivariate interpolation of arbitrarily spaced data by moving least squares methods*, J. Comput. Appl. Math. **16** 79-93. MR **87j:**65013

[Fr]   R. Franke, 1982 *Scattered data interpolation: Tests of some methods*, Math. Comp. **38**, No. 157, 181-200. MR **82m:**65008

[FrNi] R. Franke and G. Nielson, 1980 *Smooth interpolation of large sets of scattered data*, Internat. J. Numer. Methods Engrg. **15** 1691-1704. MR **82d:**65011

[LS]   P. Lancaster and K. Salkauskas, 1981 *Surfaces generated by moving least squares methods*, Math. Comp. **37**, no. 155, 141-158. MR **83c:**65015

[Mc1]   D. H. McLain, 1974 *Drawing contours from arbitrary data points*, Comput. J. **17** 318-324.

[Mc2]   D. H. McLain, 1976 *Two dimensional interpolation from random data*, Comput. J. **19** 178-181. MR **55:**4601

[Sh]   D. Shepard, 1968 *A two dimensional interpolation function for irregularly spaced data*, Proc. 23th Nat. Conf. ACM, 517-523.

School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
E-mail address: `levin@math.tau.ac.il`