Chapter 7

Moving Least Squares Approximation

An alternative to radial basis function interpolation and approximation is the so-called *moving least squares* method. As we will see below, in this method the approximation $\mathcal{P}f$ to f is obtained by solving many (small) linear systems, instead of via solution of a single – but large – linear system as we did in the previous chapters.

To make a connection with the previous chapters we start with the Backus-Gilbert formulation of the moving least squares method since this corresponds to a linearly constrained quadratic minimization problem.

7.1 Moving Least Squares Approximation: The Backus-Gilbert Approach

The connection between the standard moving least squares formulation (to be explained in the next section) and Backus-Gilbert theory was pointed out by Bos and Šalkauskas in [67]. Mathematically, in the Backus-Gilbert approach one considers a quasi-interpolant of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}), \tag{7.1}$$

where $\boldsymbol{f} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$ represents the given data. From Theorem 5.5.6 we know that the quasi-interpolant that minimizes the point-wise error is given if the *generating functions* Ψ_i are cardinal functions, i.e., $\Psi_i(\boldsymbol{x}_j) = \delta_{ij}$, $i, j = 1, \ldots, N$.

In the moving least squares method one does not attempt to minimize the pointwise error, but instead seeks to find the values of the generating functions $\Psi_i(\boldsymbol{x}) = \Psi(\boldsymbol{x}, \boldsymbol{x}_i)$ by minimizing

$$\frac{1}{2}\sum_{i=1}^{N} \Psi_i^2(\boldsymbol{x})\frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_i)} \tag{7.2}$$

subject to the polynomial reproduction constraints

$$\sum_{i=1}^{N} p(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{for all } p \in \Pi_d^s, \tag{7.3}$$

where Π_d^s is the space of s-variate polynomials of total degree at most d which has dimension $m = \binom{s+d}{d}$.

Remarks:

1. In the above formulation there is no explicit emphasis on nearness of fit as this is implicitly obtained by the quasi-interpolation "ansatz" and the closeness of the generating functions to the pointwise optimal delta functions. This is achieved by the above problem formulation if the $W(\cdot, \boldsymbol{x}_i)$ are weight functions that decrease with distance from the origin. Many of the radial functions used earlier are candidates for the weight functions. However, strict positive definiteness is not required, so that, e.g., (radial or tensor product) B-splines can also be used. The polynomial reproduction constraint is added so that the quasi-interpolant will achieve a desired approximation order. This will become clear in Section 7.6 below.

2. The smoothness functional (7.2) used here is also motivated by practical applications. In the Backus-Gilbert theory which was developed in the context of geophysics (see [17]) it is desired that the generating functions Ψ_i are as close as possible to the ideal cardinal functions (i.e., delta functions). Therefore, one needs to minimize their "spread". The polynomial reproduction constraints correspond to *discrete moment conditions* for the function $\Psi = \Psi(\boldsymbol{x}, \cdot)$.

If we think of $\boldsymbol{x}$ as a fixed (evaluation) point, then we have another constrained quadratic minimization problem of the form discussed in previous chapters. The unknowns are collected in the "coefficient vector" $\Psi(\boldsymbol{x}) = [\Psi(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, \Psi(\boldsymbol{x}, \boldsymbol{x}_N)]^T$. The smoothness functional (7.2)

$$\frac{1}{2}\Psi(\boldsymbol{x})^T Q(\boldsymbol{x})\Psi(\boldsymbol{x})$$

is given via the diagonal matrix

$$Q(\boldsymbol{x}) = \operatorname{diag}\left(\frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_1)}, \ldots, \frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_N)}\right), \tag{7.4}$$

where $W(\cdot, \boldsymbol{x}_i)$ are positive *weight functions* (and thus for any $\boldsymbol{x}$ the matrix $Q(\boldsymbol{x})$ is positive definite).

The linear polynomial reproduction constraint (7.3) can be written in matrix form as

$$A\Psi(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}),$$

where A is the $m \times N$ matrix with entries $A_{ji} = p_j(\boldsymbol{x}_i)$, $i = 1, \ldots, N$, $j = 1, \ldots, m$, and $\boldsymbol{p} = [p_1, \ldots, p_m]^T$ is a vector that contains a basis for the space Π_d^s of polynomials of degree d.

According to our earlier work we use Lagrange multipliers and then know that (cf. (6.4) and (6.5))

$$\boldsymbol{\lambda}(\boldsymbol{x}) = \left(AQ^{-1}(\boldsymbol{x})A^T\right)^{-1}\boldsymbol{p}(\boldsymbol{x}) \tag{7.5}$$

$$\Psi(\boldsymbol{x}) = Q^{-1}(\boldsymbol{x})A^T\boldsymbol{\lambda}(\boldsymbol{x}). \tag{7.6}$$

Equation (7.5) implies that the Lagrange multipliers are obtained as the solution of a Gram system

$$G(\boldsymbol{x})\boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}),$$

where the entries of G are the weighted ℓ_2 inner products

$$G_{jk}(\boldsymbol{x}) = \langle p_j, p_k \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} p_j(\boldsymbol{x}_i) p_k(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i), \quad j,k = 1, \ldots, m. \tag{7.7}$$

The special feature here is that the weight varies with the evaluation point $\boldsymbol{x}$.

Two short comments are called for. First, the Gram matrix is symmetric and positive definite since the polynomial basis is linearly independent and the weights are positive. Second, in practice, the polynomials will be represented in shifted form, i.e., centered at the point of evaluation $\boldsymbol{x}$, so that only $p_1(\boldsymbol{x}) \equiv 1 \neq 0$.

Equation (7.6) can be written componentwise, i.e., the generating functions in (7.1) are given by

$$\Psi_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i) \sum_{j=1}^{m} \lambda_j(\boldsymbol{x}) p_j(\boldsymbol{x}_i), \qquad i = 1, \ldots, N.$$

Therefore, once the values of the Lagrange multipliers $\lambda_j(\boldsymbol{x})$, $j = 1, \ldots, N$, have been determined we have explicit formulas for the values of the generating functions. In general, however, finding the Lagrange multipliers involves solving a (small) linear system that changes as soon as $\boldsymbol{x}$ changes.

7.2 Standard Interpretation of MLS Approximation

We now consider the following approximation problem. Assume we are given data values $f(\boldsymbol{x}_i)$, $i = 1, \ldots, N$, on some set $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^s$ of distinct data sites, where f is some (smooth) function, as well as an approximation space $\mathcal{U} = \text{span}\{u_1, \ldots, u_m\}$ (with $m < N$), along with the same weighted ℓ_2 inner product

$$\langle f, g \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_i) W_i(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s \text{ fixed}, \tag{7.8}$$

as introduced above in (7.7). Again, the positive weights W_i, $i = 1, \ldots, N$, depend on the evaluation point $\boldsymbol{x}$. We will interpret the weight functions in a way similar to radial basis functions, i.e., $W_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)$, with the points $\boldsymbol{x}_i$ coming from the set $\mathcal{X}$.

We now wish to find the best approximation from $\mathcal{U}$ to f at the point $\boldsymbol{x}$ with respect to the norm induced by (7.8). This means we will obtain the approximation (at the point $\boldsymbol{x}$) as

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x}) u_j(\boldsymbol{x}), \tag{7.9}$$

where the coefficients $c_j(\boldsymbol{x})$ are such that

$$\sum_{i=1}^{N} \left[\mathcal{P}f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i)\right]^2 W_i(\boldsymbol{x}) \tag{7.10}$$

is minimized. Due to the definition of the inner product (7.8) whose weight function "moves" with the evaluation point $\boldsymbol{x}$, the coefficients c_j in (7.9) depend also on $\boldsymbol{x}$. As a consequence one has to solve the normal equations

$$\sum_{j=1}^{m} c_j(\boldsymbol{x})\langle u_j, u_k\rangle_{W(\boldsymbol{x})} = \langle f, u_k\rangle_{W(\boldsymbol{x})}, \qquad k = 1, \ldots, m, \tag{7.11}$$

anew each time the evaluation point $\boldsymbol{x}$ is changed. In matrix notation (7.11) becomes

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_u(\boldsymbol{x}), \tag{7.12}$$

with the positive definite Gram matrix $G(\boldsymbol{x}) = \left(\langle u_j, u_k\rangle_{W(\boldsymbol{x})}\right)_{j,k=1}^{m}$, coefficient vector $\boldsymbol{c}(\boldsymbol{x})$ and right-hand side vector $\boldsymbol{f}_u(\boldsymbol{x})$ as in (7.11) all depending on $\boldsymbol{x}$.

In the moving least squares method one usually takes $\mathcal{U}$ to be a space of (multivariate) polynomials, i.e.,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x})p_j(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{7.13}$$

where the $\{p_1, \ldots, p_m\}$ is a basis for the space Π_d^s of s-variate polynomials of degree d.

The Gram system (7.12) now becomes

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x}), \tag{7.14}$$

where the matrix $G(\boldsymbol{x})$ has entries

$$G_{jk}(\boldsymbol{x}) = \langle p_j, p_k\rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} p_j(\boldsymbol{x}_i)p_k(\boldsymbol{x}_i)W(\boldsymbol{x}, \boldsymbol{x}_i), \quad j, k = 1, \ldots, m, \tag{7.15}$$

and the right-hand side vector consists of the projections of the data f onto the basis functions, i.e.,

$$\boldsymbol{f}_p(\boldsymbol{x}) = \left[\langle f, p_1\rangle_{W(\boldsymbol{x})}, \ldots, \langle f, p_m\rangle_{W(\boldsymbol{x})}\right]^T.$$

Remarks:

1. The fact that the coefficients depend on the evaluation point $\boldsymbol{x}$, and thus for every evaluation of $\mathcal{P}f$ a Gram system (with different matrix $G(\boldsymbol{x})$) needs to be solved, initially scared people away from the moving least squares approach. However, one can either choose compactly supported weight functions so that only a few terms are "active" in the sum in (7.15), or even completely avoid the solution of linear systems (see, e.g., [202]).

2. We point out that since we are working with a polynomial basis, the matrix G can also be interpreted as a moment matrix for the weight W. This interpretation is used in the engineering literature (see, e.g., [381]), and also plays an essential role when connecting moving least squares approximation to the more efficient concept of *approximate approximation* [434]. For a discussion of approximate moving least squares approximation see [203, 204, 205, 206].

The connection to the constrained quadratic minimization problems discussed earlier can be seen as follows. We are now minimizing (for fixed $\boldsymbol{x}$)

$$\frac{1}{2}\boldsymbol{c}^T(\boldsymbol{x})G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{\mu}^T(\boldsymbol{x})\left[G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) - AQ^{-1}(\boldsymbol{x})\boldsymbol{f}\right], \tag{7.16}$$

where $G(\boldsymbol{x})$ is the Gram matrix (7.7), $Q(\boldsymbol{x})$ the diagonal matrix of weight functions (7.4) and A the matrix of polynomials used earlier. The term $AQ^{-1}(\boldsymbol{x})\boldsymbol{f}$ corresponds to the right-hand side vector $\boldsymbol{f}_p(\boldsymbol{x})$ of (7.14). The solution of the linear system resulting from the minimization problem (7.16) gives us

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{x}) &= \left(G(\boldsymbol{x})G^{-1}(\boldsymbol{x})G^T(\boldsymbol{x})\right)^{-1} AQ^{-1}(\boldsymbol{x})\boldsymbol{f} = G^{-T}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f}\\ \boldsymbol{c}(\boldsymbol{x}) &= G^{-1}(\boldsymbol{x})G^T(\boldsymbol{x})\boldsymbol{\mu}(\boldsymbol{x}) = \boldsymbol{\mu}(\boldsymbol{x})\end{aligned}$$

so that – as in the case of radial basis function interpolation – by solving only the Gram system $G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x})$ we automatically minimize the functional

$$\begin{aligned}\boldsymbol{c}^T(\boldsymbol{x})G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) &= \sum_{j=1}^{m}\sum_{k=1}^{m} c_j(\boldsymbol{x})c_k(\boldsymbol{x})G_{jk}(\boldsymbol{x})\\ &= \sum_{j=1}^{m}\sum_{k=1}^{m} c_j(\boldsymbol{x})c_k(\boldsymbol{x})\langle p_j, p_k\rangle_{W(\boldsymbol{x})}\end{aligned}$$

which should be interpreted as the native space norm of the approximant $\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x})p_j(\boldsymbol{x})$.

7.3 A Dual Representation for the Standard Approach

We now know that on the one hand (from the Backus-Gilbert formulation)

$$G(\boldsymbol{x})\boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}) \quad \Longleftrightarrow \quad \boldsymbol{\lambda}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x}). \tag{7.17}$$

By taking multiple right-hand sides $\boldsymbol{p}(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathcal{X}$ we get

$$\Lambda = G^{-1}(\boldsymbol{x})A, \tag{7.18}$$

where the $m \times N$ matrix Λ has entries $\Lambda_{ji} = \lambda_j(\boldsymbol{x}_i)$.

The standard formulation, on the other hand, gives us

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x}) \quad \Longleftrightarrow \quad \boldsymbol{c}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})\boldsymbol{f}_p(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f} \tag{7.19}$$

with

$$\boldsymbol{f}_p(\boldsymbol{x}) = \left[\langle f, p_1\rangle_{W(\boldsymbol{x})}, \ldots, \langle f, p_m\rangle_{W(\boldsymbol{x})}\right]^T = AQ^{-1}(\boldsymbol{x})\boldsymbol{f}$$

as above. By combining (7.18) with (7.19) we get

$$\boldsymbol{c}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f} = \Lambda Q^{-1}(\boldsymbol{x})\boldsymbol{f} = \boldsymbol{f}_\lambda(\boldsymbol{x}),$$

where $\boldsymbol{f}_\lambda(\boldsymbol{x})$ is defined analogously to $\boldsymbol{f}_p(\boldsymbol{x})$. Componentwise this gives us

$$c_j(\boldsymbol{x}) = \langle f, \lambda_j \rangle_{W(\boldsymbol{x})}, \qquad j = 1, \ldots, m,$$

and therefore,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} \langle f, \lambda_j \rangle_{W(\boldsymbol{x})} p_j(\boldsymbol{x}). \tag{7.20}$$

It is also possible to formulate the moving least squares method by using the Lagrange multipliers of the Backus-Gilbert approach as basis functions for the approximation space $\mathcal{U}$. Then, using the same argumentation as above, we end up with

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} d_j(\boldsymbol{x}) \lambda_j(\boldsymbol{x}) \tag{7.21}$$

with

$$d_j(\boldsymbol{x}) = \langle f, p_j \rangle_{W(\boldsymbol{x})}, \qquad j = 1, \ldots, m.$$

Remarks:

1. The Lagrange multipliers form a basis that is *dual* to the polynomials. In particular one can show that for any $\boldsymbol{x} \in \mathcal{X}$

   $$\langle \lambda_j, p_k \rangle_{W(\boldsymbol{x})} = \delta_{jk}, \qquad j, k = 1, \ldots, m.$$

   This shows that we have two sets of basis functions that are *bi-orthogonal* on the set $\mathcal{X}$.

2. Note that the expansions (7.20) and (7.21) are generalizations of (finite) eigenfunction or Fourier series expansions.

7.4 Equivalence of Our Approaches to Moving Least Squares Approximation

We now show that the two main approaches to the moving least squares method described above are equivalent, i.e., we show that $\mathcal{P}f(\boldsymbol{x})$ computed via (7.1) and (7.13) are the same. The approximant (7.1) in the Backus-Gilbert "ansatz" is of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i) \Psi_i(\boldsymbol{x}) = \Psi^T(\boldsymbol{x}) \boldsymbol{f},$$

where as before $\Psi(\boldsymbol{x}) = [\Psi(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, \Psi(\boldsymbol{x}, \boldsymbol{x}_N)]^T$ and $\boldsymbol{f} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$. The standard moving least squares formulation (7.13), on the other hand, establishes $\mathcal{P}f(\boldsymbol{x})$ in the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x}) p_j(\boldsymbol{x}) = \boldsymbol{p}^T(\boldsymbol{x}) \boldsymbol{c}(\boldsymbol{x}),$$

where $\boldsymbol{p} = [p_1, \ldots, p_m]^T$ and $\boldsymbol{c}(\boldsymbol{x}) = [c_1(\boldsymbol{x}), \ldots, c_m(\boldsymbol{x})]^T$.

In (7.16) we wrote the normal equations for the standard approach as

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = AQ^{-1}(\boldsymbol{x})\boldsymbol{f}$$

which implies

$$\boldsymbol{c}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f}.$$

Thus, using the standard approach, we get

$$\mathcal{P}f(\boldsymbol{x}) = \boldsymbol{p}^T(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{p}^T(\boldsymbol{x})G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f}. \tag{7.22}$$

For the Backus-Gilbert approach we derived (see (7.5) and (7.6))

$$\begin{aligned} \boldsymbol{\lambda}(\boldsymbol{x}) &= G^{-1}(\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x}) \\ \Psi(\boldsymbol{x}) &= Q^{-1}(\boldsymbol{x})A^T\boldsymbol{\lambda}(\boldsymbol{x}), \end{aligned}$$

where $G(\boldsymbol{x}) = AQ^{-1}(\boldsymbol{x})A^T$ (see (7.7) or (7.15)). Therefore, we now obtain

$$\mathcal{P}f(\boldsymbol{x}) = \Psi^T(\boldsymbol{x})\boldsymbol{f} = \left[Q^{-1}(\boldsymbol{x})A^TG^{-1}(\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x})\right]^T\boldsymbol{f}$$

which, by the symmetry of $Q(\boldsymbol{x})$ and $G(\boldsymbol{x})$, is the same as (7.22).

Remarks:

1. The equivalence of the two approaches shows that the moving least squares approximant has all of the following properties:

   - It reproduces any polynomial of degree at most d in s variables exactly.
   - It produces the best locally weighted least squares fit.
   - Viewed as a quasi-interpolant, the generating functions Ψ_i are as close as possible to the optimal cardinal basis functions in the sense that (7.2) is minimized.
   - Since polynomials are infinitely smooth, either of the representations of $\mathcal{P}f$ shows that its smoothness is determined by the smoothness of the weight function(s) $W_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)$.

2. In particular, the standard moving least squares method will reproduce the polynomial basis functions $p_1, \ldots, p_m$ even though this is not explicitly enforced by the minimization (solution of the normal equations). Moreover, the more general "ansatz" with approximation space $\mathcal{U}$ allows us to build moving least squares approximations that also reproduce any other function that is included in $\mathcal{U}$. This can be very beneficial for the solution of partial differential equations with known singularities (see, e.g., the papers [16] by Babuška and Melenk, and [49] by Belytschko and co-authors).

By also considering the dual expansion (7.21) we have three alternative representations for the moving least squares quasi-interpolant. This is summarized in the following theorem.

Theorem 7.4.1 *Let $f : \Omega \to \mathbb{R}$ be some function whose values on the set of points $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^N \subset \mathbb{R}^s$ are given as data. Let $p_1, \ldots, p_m$ be a basis for Π_d^s, let $\{W(\cdot, \boldsymbol{x}_i)\}_{i=1}^N$ be a set of positive weight functions centered at the points of $\mathcal{X}$, and let λ_j, $j = 1, \ldots, m$, be the Lagrange multipliers defined by (7.5). Furthermore, consider the generating functions*

$$\Psi_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i) \sum_{j=1}^m \lambda_j(\boldsymbol{x}) p_j(\boldsymbol{x}_i), \qquad i = 1, \ldots, N.$$

The best local least squares approximation to f on $\mathcal{X}$ in the sense of (7.10) is given by

$$\begin{aligned} \mathcal{P}f(\boldsymbol{x}) &= \sum_{j=1}^m \langle f, \lambda_j \rangle_{W(\boldsymbol{x})} p_j(\boldsymbol{x}) \\ &= \sum_{j=1}^m \langle f, p_j \rangle_{W(\boldsymbol{x})} \lambda_j(\boldsymbol{x}) \\ &= \sum_{i=1}^N f(\boldsymbol{x}_i) \Psi_i(\boldsymbol{x}). \end{aligned}$$

7.5 Examples

7.5.1 Shepard's Method

The moving least squares method in the case $m = 1$ with $p_1(\boldsymbol{x}) \equiv 1$ is known to yield Shepard's method [578]. In the statistics literature Shepard's method is known as a *kernel method* (see, e.g., the papers from the 1950s and 60s [534, 501, 476, 623]). Using our notation we have

$$\mathcal{P}f(\boldsymbol{x}) = c_1(\boldsymbol{x}).$$

The Gram "matrix" consists of only one element

$$G(\boldsymbol{x}) = \langle p_1, p_1 \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^N W(\boldsymbol{x}, \boldsymbol{x}_i)$$

so that

$$G(\boldsymbol{x}) \boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x})$$

implies

$$c_1(\boldsymbol{x}) = \frac{\displaystyle\sum_{i=1}^N f(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i)}{\displaystyle\sum_{i=1}^N W(\boldsymbol{x}, \boldsymbol{x}_i)}.$$

The dual basis is defined by

$$G(\boldsymbol{x}) \boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x})$$

so that

$$\lambda_1(\boldsymbol{x}) = \frac{1}{\sum_{i=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_i)},$$

and

$$\mathcal{P}f(\boldsymbol{x}) = d_1(\boldsymbol{x})\lambda_1(\boldsymbol{x}) \tag{7.23}$$

with

$$d_1(\boldsymbol{x}) = \langle f, p_1 \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} f(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i).$$

The generating functions are defined as

$$\Psi_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)\lambda_1(\boldsymbol{x})p_1(\boldsymbol{x}_i) = \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{i=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_i)}.$$

This gives rise to the well-known quasi-interpolation formula for Shepard's method

$$\begin{aligned} \mathcal{P}f(\boldsymbol{x}) &= \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) \\ &= \sum_{i=1}^{N} f(\boldsymbol{x}_i) \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{k=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_k)}. \end{aligned}$$

Of course this is the same as the basis expansion $c_1(\boldsymbol{x})$ and the dual expansion (7.23).

We should now have bi-orthogonality of the basis and dual basis, i.e.,

$$\langle \lambda_1, p_1 \rangle_{W(\boldsymbol{x})} = 1.$$

Indeed

$$\begin{aligned} \langle \lambda_1, p_1 \rangle_{W(\boldsymbol{x})} &= \sum_{i=1}^{m} \lambda_1(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i) \\ &= \sum_{i=1}^{N} \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{k=1}^{N} W(\boldsymbol{x}_i, \boldsymbol{x}_k)}, \end{aligned}$$

and this equals 1 if we restrict $\boldsymbol{x}$ to be an element of the set $\mathcal{X}$.

7.5.2 Plots of Basis-Dual Basis Pairs

We also illustrate the moving least squares basis functions, dual basis functions and generating functions for a one-dimensional example with $\mathcal{X}$ being the set of 13 equally spaced points in $[-5, 5]$. We take $m = 2$, i.e., we consider the case that ensures

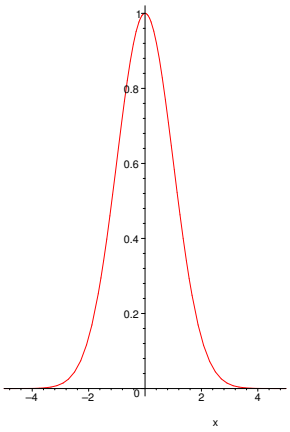

Figure 7.1: Plot of Gaussian weight function centered at $\boldsymbol{x}_7 = 0$.

reproduction of quadratic polynomials. The weight function is taken to be a standard Gaussian as depicted in Figure 7.1.

The three basis polynomials $p_1(x) = 1$, $p_2(x) = x$, and $p_3(x) = x^2$ are shown in Figure 7.2, whereas the dual basis functions λ_1, λ_2, and λ_3 are displayed in Figure 7.3. The figure shows that, except for the boundary effects caused by the finite interval, these functions resemble a quadratic, linear and constant polynomial.

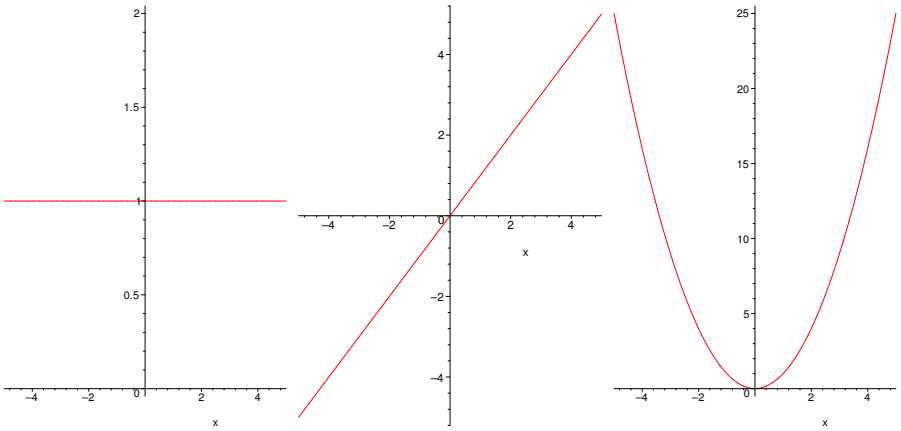

Figure 7.2: Plot of three polynomial basis functions for moving least squares approximation.

In Figure 7.4 we plot one of the generating functions (centered at $\boldsymbol{x}_7 = 0$) along with an approximate moving least squares generating function of the form

$$\Psi(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\sqrt{\sigma\pi}} \left(\frac{3}{2} - \frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{\sigma} \right) e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{\sigma}}$$

with scale parameter σ as derived in [205].

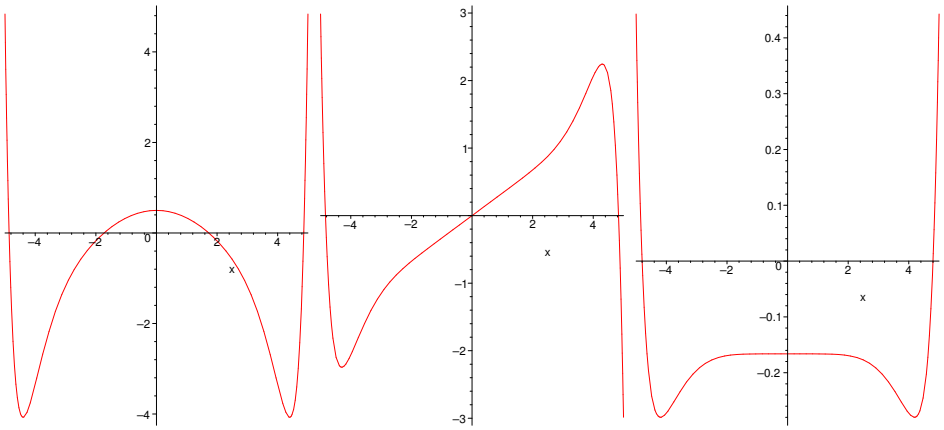

Figure 7.3: Plot of three dual basis functions for moving least squares approximation.

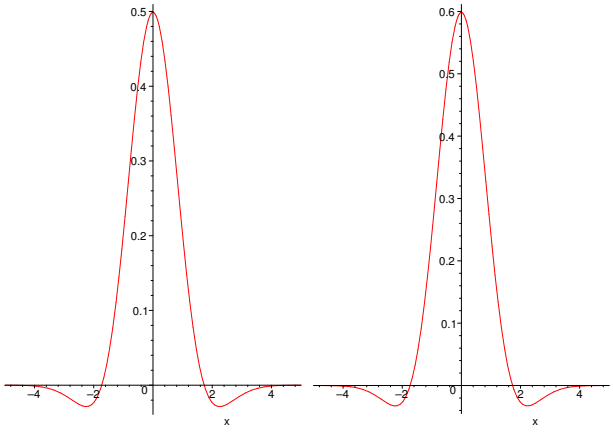

Figure 7.4: Plot of moving least squares generating function (left) and approximate generating function (right) centered at $\boldsymbol{x}_7 = 0$.

7.6 Approximation Order of Moving Least Squares

Since the moving least squares approximants can be written as quasi-interpolants, we can use standard techniques to derive their point-wise error estimates. The standard argument proceeds as follows. Let f be a given (smooth) function that generates the data, i.e., $f_1 = f(\boldsymbol{x}_1), \ldots, f_N = f(\boldsymbol{x}_N)$, and let p be an arbitrary polynomial. Moreover, assume that the moving least squares approximant is given in the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x})$$

with the generating functions Ψ_i satisfying the polynomial reproduction property

$$\sum_{i=1}^{N} p(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{for all } p \in \Pi_d^s,$$

as described at the beginning of this chapter. Then, due to the polynomial reproduction property of the generating functions,

$$\begin{aligned} |f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| &\leq |f(\boldsymbol{x}) - p(\boldsymbol{x})| + |p(\boldsymbol{x}) - \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x})| \\ &= |f(\boldsymbol{x}) - p(\boldsymbol{x})| + |\sum_{i=1}^{N} p(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) - \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x})| \\ &\leq |f(\boldsymbol{x}) - p(\boldsymbol{x})| + \sum_{i=1}^{N} |p(\boldsymbol{x}_i) - f(\boldsymbol{x}_i)||\Psi_i(\boldsymbol{x})| \\ &\leq \|f - p\|_\infty \left[1 + \sum_{i=1}^{N} |\Psi_i(\boldsymbol{x})|\right]. \end{aligned} \tag{7.24}$$

We see that in order to refine the error estimate we now have to answer two questions:

- How well do polynomials approximate f? This will be done with standard Taylor expansions.
- Are the generating functions bounded? The expression $\sum_{i=1}^{N} |\Psi_i(\boldsymbol{x})|$ is known as the *Lebesgue function*, and finding a bound for the Lebesgue function is the main task that remains.

By taking the polynomial p above to be the Taylor polynomial for f at $\boldsymbol{x}$ of total degree d, the remainder term immediately yields an estimate of the form

$$\begin{aligned} \|f - p\|_\infty &\leq C_1 h^{d+1} \max_{\boldsymbol{x}\in\Omega} |D^{\boldsymbol{\alpha}} f(\boldsymbol{x})|, \qquad |\boldsymbol{\alpha}| = d+1, \\ &= C_1 h^{d+1} |f|_{d+1}, \end{aligned} \tag{7.25}$$

where we have used the abbreviation

$$|f|_{d+1} = \max_{\boldsymbol{x}\in\Omega} |D^{\boldsymbol{\alpha}} f(\boldsymbol{x})|, \qquad |\boldsymbol{\alpha}| = d+1.$$

Thus, if we can establish a uniform bound for the Lebesgue function, then (7.24) and (7.25) will result in

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \leq C h^{d+1} |f|_{d+1}$$

which shows that moving least squares approximation with polynomial reproduction of degree d has approximation order $\mathcal{O}(h^{d+1})$.

For Shepard's method, i.e., moving least squares approximation with constant reproduction (i.e., $m = 1$ or $d = 0$), we saw above that the generating functions are of the form

$$\Psi_i(\boldsymbol{x}) = \frac{W(\boldsymbol{x}, \boldsymbol{x}_i)}{\sum_{j=1}^{N} W(\boldsymbol{x}, \boldsymbol{x}_j)}$$

and therefore the Lebesgue function admits the uniform bound

$$\sum_{i=1}^{N} |\Psi_i(\boldsymbol{x})| = 1,$$

This shows that Shepard's method has approximation order $\mathcal{O}(h)$.

Bounding the Lebesgue function in the general case is more involved and is the subject of the papers [378] by Levin and [632] by Wendland. This results in approximation order $\mathcal{O}(h^{d+1})$ for a moving least squares method that reproduces polynomials of degree d. In both papers the authors assumed that the weight function is compactly supported, and that the support size is scaled proportional to the fill distance. However, similar estimates should be possible if the weight function only decays fast enough (see, e.g., the survey by de Boor [60]).

Aside from this consideration, the choice of weight function W does not play a role in determining the approximation order of the moving least squares method. As noted earlier, it only determines the smoothness of $\mathcal{P}f$. For example, in the paper [146] from the statistics literature on local regression the authors state that often "the choice [of weight function] is not too critical", and the use of the so-called *tri-cube*

$$W(\boldsymbol{x}, \boldsymbol{x}_i) = (1 - \|\boldsymbol{x} - \boldsymbol{x}_i\|^3)_+^3, \qquad \boldsymbol{x} \in \mathbb{R}^s,$$

is suggested. Of course, many other weight functions such as (radial) B-splines or any of the (bell-shaped) radial basis functions studied earlier can also be used. If the weight function is compactly supported, then the generating functions Ψ_i will be so, too. This leads to computationally efficient methods since the Gram matrix $G(\boldsymbol{x})$ will be sparse.

An interesting question is also the size of the support of the different local weight functions. Obviously, a fixed support size for all weight functions is possible. However, this will cause serious problems as soon as the data are not uniformly distributed. Therefore, in the arguments in [378] and [632] the assumption is made that the data are at least quasi-uniformly distributed. Another choice for the support size of the individual weight functions is based on the number of nearest neighbors, i.e., the support size is chosen so that each of the local weight functions contains the same number of centers in its support. A third possibility is suggested by Schaback [556]. He proposes to use another moving least squares approximation based on (equally spaced) auxiliary points to determine a smooth function δ so that at each evaluation point $\boldsymbol{x}$ the radius of the support of the weight function is given by $\delta(\boldsymbol{x})$. However, convergence estimates for these latter two choices do not exist.

Sobolev error estimates are provided for moving least squares approximation with compactly supported weight functions in [7]. The rates obtained in that paper are not in terms of the fill distance but instead in terms of the support size R of the weight function. Moreover, it is assumed that for general s and $m = \binom{s+d}{d}$ the local Lagrange functions are bounded. As mentioned above, this is the hard part, and such bounds are only provided in the case $s = 2$ with $d = 1$ and $d = 2$ in [7]. However, if combined with the general bounds for the Lebesgue function provided by Wendland the paper [7] yields the following estimates:

$$|f(\boldsymbol{x}) - \mathcal{P}f(\boldsymbol{x})| \leq CR^{d+1}|f|_{d+1}$$

but also

$$|\nabla(f - \mathcal{P}f)(\boldsymbol{x})| \leq CR^d |f|_{d+1}.$$

In the weaker L_2-norm we have

$$\|f - \mathcal{P}f\|_{L_2(B_j \cap \Omega)} \leq CR^{d+1} |f|_{W_2^{d+1}(B_j \cap \Omega)}$$

and

$$\|\nabla(f - \mathcal{P}f)\|_{L_2(B_j \cap \Omega)} \leq CR^d |f|_{W_2^{d+1}(B_j \cap \Omega)},$$

where the balls B_j provide a finite cover of the domain Ω, i.e., $\Omega \subseteq \bigcup_j B_j$, and the number of overlapping balls is bounded.

Remarks:

1. In the statistics literature the moving least squares idea is known as local (polynomial) regression. There is a book by Fan and Gijbels [186] and a review article by Cleveland and Loader [146] according to which the basic ideas of local regression can be traced back at least to work of Gram [267], Woolhouse [648], and De Forest [148, 149] from the 1870s and 1880s.

2. In particular, in the statistics literature one learns that the use of least squares approximation is justified when the data $f_1, \ldots, f_N$ are normally distributed, whereas, if the noise in the data is not Gaussian, then other criteria should be used. See, e.g., the survey article [146] for more details.

3. The general moving least squares method first appeared in the approximation theory literature in a paper by Lancaster and Šalkauskas [358] who also pointed out the connection to earlier (more specialized) work by Shepard [578] and McLain [436].

4. Early error estimates for some special cases were provided by Farwig in [188, 189].