Dual Bases and Discrete Reproducing Kernels: A Unified Framework for RBF and MLS Approximation

G. E. Fasshauer*

Abstract

Moving least squares (MLS) and radial basis function (RBF) methods play a central role in multivariate approximation theory. In this paper we provide a unified framework for both RBF and MLS approximation. This framework turns out to be a linearly constrained quadratic minimization problem. We show that RBF approximation can be considered as a special case of MLS approximation. This sheds new light on both MLS and RBF approximation. Among the new insights are dual bases for the approximation spaces and certain discrete reproducing kernels.

1 Introduction

In the literature on multivariate approximation methods (see, e.g., the recent books [1, 5, 6, 17] or the references therein) both the moving least squares (MLS) method and radial basis functions (RBFs) are prominently featured. It seems that authors from the engineering community prefer methods for the solution of partial differential equations that are based on the moving least squares approach (such as EFG, *hp*-clouds, MLPG, PUFEM, RKPM, SPH or others), whereas the mathematics community has focussed mostly on radial basis functions. In this paper we will consider the approximation (or interpolation) of scattered multivariate data of the form $(\boldsymbol{x}_i, f_i)$, $i = 1, \ldots, N$, where the data sites $\boldsymbol{x}_i$ are arbitrarily distributed in $\mathbb{R}^s$ and the associated values $f_i = f(\boldsymbol{x}_i) \in \mathbb{R}$ are assumed to be generated by some (unknown) function f. In particular, we will discuss the use of both RBFs and MLS approximation to accomplish this task. It is our goal to show that, taking an appropriate point of view (namely that of a linearly constrained quadratic optimization problem), the discussion of the two approaches can be unified.

The connection between radial basis function interpolation and optimal recovery has been known for some time (see, e.g., [19] or the recent paper [22]). It is known that every strictly positive definite (and with some extra effort also conditionally positive definite) kernel $\Phi(\cdot, \cdot)$ can be associated with a natural Hilbert function space – its *native space* $\mathcal{N}_\Phi$. In the case of strictly positive definite kernels, this native space is a reproducing kernel Hilbert space with Φ as its reproducing kernel. The most important consequence of this is the connection to optimal recovery, i.e., the fact that the solution of the scattered data interpolation problem to data generated by some function f from

*Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, USA, fass@amadeus.math.iit.edu

the native space $\mathcal{N}_\Phi$ using linear combinations of the basis functions $\Phi_i = \Phi(\cdot, \boldsymbol{x}_i)$ yields not only an interpolant for the data, but an *optimal* one. In fact, we can state this optimality in (at least) three different ways (see, e.g., [20, 21] or [30]):

1. The interpolant has minimal native space norm.
2. The interpolant is the (native space) best approximation to the data.
3. If the interpolant is written in cardinal form, then it has the smallest pointwise error among all possible quasi-interpolants.

In particular, the first of these statements implies that the RBF interpolant is in fact the (automatic) solution to a constrained quadratic optimization problem. We will now elaborate on this point of view. However, we are interested in the more general setting where we still sample the function f on the set $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, but now introduce a second set $\Xi = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_M\}$ at which we center the basis functions. Usually we will have $M \leq N$, and we will see below that the case $M = N$ with $\Xi = \mathcal{X}$ recovers the traditional interpolation setting.

Let us assume we want to use a function $\mathcal{P}f$ (RBF expansion) of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \Phi(\boldsymbol{x}, \boldsymbol{\xi}_j), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{1}$$

where (for the sake of simplicity) Φ is assumed to be a strictly positive definite kernel such as a Gaussian $\Phi(\boldsymbol{x}, \boldsymbol{\xi}_j) = e^{-\alpha \|\boldsymbol{x} - \boldsymbol{\xi}_j\|^2}$, $\alpha > 0$, an inverse multiquadric $\Phi(\boldsymbol{x}, \boldsymbol{\xi}_j) = (\|\boldsymbol{x} - \boldsymbol{\xi}_j\|^2 + \alpha^2)^{-1/2}$, or one of the compactly supported functions suggested by Wendland, e.g., $\Phi(\boldsymbol{x}, \boldsymbol{\xi}_j) = (1 + \|\boldsymbol{x} - \boldsymbol{\xi}_j\|)_+^4 (4\|\boldsymbol{x} - \boldsymbol{\xi}_j\| + 1)$.

We now choose the coefficients c_j in (1) such that the quadratic form

$$\frac{1}{2} \boldsymbol{c}^T Q \boldsymbol{c} \tag{2}$$

with $\boldsymbol{c} = [c_1, \ldots, c_M]^T$ and some symmetric positive definite matrix Q is minimized subject to the N linear constraints

$$A\boldsymbol{c} = \boldsymbol{f}, \tag{3}$$

where A is an $N \times M$ matrix assumed to have full rank, and the right-hand side $\boldsymbol{f} = [f_1, \ldots, f_N]^T$ is given. We will specify possible connections between the kernel Φ and the matrices Q and A shortly.

Such a constrained quadratic minimization problem can be converted to a system of linear equations by introducing *Lagrange multipliers* $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]^T$, i.e., we consider finding the minimum of

$$\frac{1}{2} \boldsymbol{c}^T Q \boldsymbol{c} - \boldsymbol{\lambda}^T [A\boldsymbol{c} - \boldsymbol{f}] \tag{4}$$

with respect to $\boldsymbol{c}$ and $\boldsymbol{\lambda}$. Since Q is a positive definite matrix, it is well known that the functional to be minimized is convex, and thus has a unique minimum. Therefore, the standard necessary condition for such a minimum (which is obtained by differentiating

with respect to $\boldsymbol{c}$ and $\boldsymbol{\lambda}$ and finding the zeros of those derivatives) is also sufficient. This leads to the linear system

$$\begin{bmatrix} Q & -A^T \\ A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{f} \end{bmatrix}.$$

By applying Gaussian elimination to this block matrix (Q is invertible since it is assumed to be positive definite and A is assumed to have full rank) we get

$$\boldsymbol{\lambda} = \left(AQ^{-1}A^T\right)^{-1} \boldsymbol{f} \tag{5}$$

$$\boldsymbol{c} = Q^{-1}A^T \left(AQ^{-1}A^T\right)^{-1} \boldsymbol{f}. \tag{6}$$

The connection to our introductory comments regarding RBF interpolation is as follows. We assume that the centers are chosen at the data locations, i.e., $M = N$ and $\Xi = \mathcal{X}$. Next, we let the quadratic form represent the native space norm (see, e.g., [20]) of the interpolant $\mathcal{P}f = \sum_{j=1}^{N} c_j \Phi(\cdot, \boldsymbol{x}_j)$, i.e.,

$$\|\mathcal{P}f\|^2_{\mathcal{N}_\Phi} = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \Phi(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{c}^T Q \boldsymbol{c} \tag{7}$$

with $Q_{ij} = \Phi(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $\boldsymbol{c} = [c_1, \ldots, c_N]^T$, and take as the linear side conditions the interpolation conditions

$$A\boldsymbol{c} = \boldsymbol{f} \qquad \Longleftrightarrow \qquad \mathcal{P}f(\boldsymbol{x}_i) = f_i, \quad i = 1, \ldots, N,$$

with the interpolation matrix A having the same entries as the (symmetric) matrix Q, the same coefficient vector $\boldsymbol{c}$ as in (7) and data vector $\boldsymbol{f} = [f_1, \ldots, f_N]^T$. Then we see from the general solution (5) and (6) that the Lagrange multipliers become

$$\boldsymbol{\lambda} = A^{-T}\boldsymbol{f} = A^{-1}\boldsymbol{f}$$

and the coefficients are given by

$$\boldsymbol{c} = \boldsymbol{\lambda}.$$

Therefore, as stated earlier, the minimum norm interpolant is obtained by solving the interpolation equations alone, i.e., the norm minimization comes for free.

Since we started out with the more general point of view that $\mathcal{P}f$ is generated by M basis functions, and N linear constraints are specified, our formulation also covers both over- and under-determined least squares fitting where the quadratic form $\boldsymbol{c}^T Q \boldsymbol{c}$ represents an added *smoothing* (or *regularization*) term.

This term is not required to obtain a unique solution of the system $A\boldsymbol{c} = \boldsymbol{f}$ in the over-determined case ($M \leq N$) which corresponds to the standard least squares approximation by an RBF expansion of the form (1). The coefficients c_j are then either determined directly by an SVD or QR solution of the linear system, or by minimizing $\|\mathcal{P}f - f\|_2^2$, where the ℓ_2-norm

$$\|f\|_2^2 = \sum_{i=1}^{N} [f(\boldsymbol{x}_i)]^2$$

is induced by the discrete inner product

$$\langle f, g \rangle = \sum_{i=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_i). \tag{8}$$

It is well known that this approximation problem has a unique solution if the (rectangular) matrix A with entries

$$A_{jk} = \Phi(\boldsymbol{x}_j, \boldsymbol{\xi}_k), \quad j = 1, \ldots, N, \ k = 1, \ldots, M,$$

has full rank. If the centers in Ξ are chosen to form a subset of the data locations $\mathcal{X}$ then A does indeed have full rank since A has an $M \times M$ square submatrix which is non-singular (by virtue of being an *interpolation matrix* and since the kernel is assumed to be strictly positive definite). Once again, by solving $A\boldsymbol{c} = \boldsymbol{f}$ (or finding the least squares RBF approximation) we are also ensured to have a minimum (native space) norm solution. More theoretical work addressing least squares radial basis function approximation was reported in [5] and [25].

In the under-determined case, however, a *regularization term* is needed to guarantee a unique solution (cf. the solution of under-determined linear systems in the numerical linear algebra literature, e.g., [26]). Usually the regularized least squares approximation problem is formulated as minimization of

$$\frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} + \omega \sum_{j=1}^{N} \left(\mathcal{P} f(\boldsymbol{x}_j) - f_j\right)^2, \tag{9}$$

where the regularization parameter ω balances the tradeoff between the quadratic form which controls the smoothness of the fitting function and the least squares term which measures the closeness to the data. The formulation (9) is used in regularization theory (see, e.g., [7, 12]). The same formulation is also used in penalized least squares fitting (see, e.g., [13]), the literature on smoothing splines [18, 23], and in papers by Wahba on thin plate splines (e.g., [28, 29]). In fact, the idea of smoothing a data fitting process by this kind of formulation seems to go back to at least Whittaker [31] in 1923. In practice a penalized least squares formulation is especially useful if the data f_i cannot be completely trusted, i.e., they are contaminated by noise. In this case, a (penalized) least squares fit is advisable. The problem of minimizing (9) is also known as ridge regression in the statistics literature, and there are many techniques (such as generalized cross-validation) for choosing a good regularization parameter ω.

The equivalence of (9) with our earlier formulation (4) follows from

$$\begin{aligned} \frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} + \omega \sum_{j=1}^{N} \left(\mathcal{P} f(\boldsymbol{x}_j) - f_j\right)^2 &= \frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} + \omega [A\boldsymbol{c} - \boldsymbol{f}]^T [A\boldsymbol{c} - \boldsymbol{f}] \\ &= \frac{1}{2}\boldsymbol{c}^T Q \boldsymbol{c} - \boldsymbol{\lambda}^T [A\boldsymbol{c} - \boldsymbol{f}], \end{aligned}$$

where

$$\boldsymbol{\lambda} = -\omega [A\boldsymbol{c} - \boldsymbol{f}],$$

and the regularization parameter is determined automatically by the native space norm minimization. This is analogous to the solution of an under-determined linear system

using the singular value decomposition, in which case one automatically obtains the minimum norm solution.

2 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 is customary for RBF approximation.

To make a direct connection with the constrained optimization formulation of the introduction we start with the Backus-Gilbert formulation of the moving least squares method.

2.1 The Backus-Gilbert Approach for MLS Approximation

In the Backus-Gilbert approach one considers a quasi-interpolant of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}), \tag{10}$$

where $\boldsymbol{f} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$ represents the given data. In the introduction we pointed out 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

$$\frac{1}{2}\sum_{i=1}^{N} \Psi_i^2(\boldsymbol{x})\frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_i)} \tag{11}$$

subject to the polynomial reproduction constraints

$$\sum_{i=1}^{N} p(\boldsymbol{x}_i)\Psi_i(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{for all } p \in \Pi_d^s, \tag{12}$$

where Π_d^s is the space of s-variate polynomials of total degree at most d which has dimension $m = \binom{s+d}{d}$ and the $W(\cdot, \boldsymbol{x}_i)$ are positive *weight functions*.

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 closeness is achieved if the weight functions decrease with distance from the origin. Many standard (strictly positive definite) radial functions are candidates for these 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 a standard requirement for quasi-interpolants used to achieve a desired approximation order.

If we think of $\boldsymbol{x}$ as a fixed (evaluation) point, then we have another constrained quadratic minimization problem of the form discussed in the introduction. 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 (11)

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

is given via the diagonal matrix

$$Q(\boldsymbol{x}) = \operatorname{diag}\left(\frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_1)}, \ldots, \frac{1}{W(\boldsymbol{x}, \boldsymbol{x}_N)}\right), \tag{13}$$

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

The linear polynomial reproduction constraint (12) 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. (5) and (6))

$$\begin{aligned} \boldsymbol{\lambda}(\boldsymbol{x}) &= \left(AQ^{-1}(\boldsymbol{x})A^T\right)^{-1}\boldsymbol{p}(\boldsymbol{x}) && (14)\\ \Psi(\boldsymbol{x}) &= Q^{-1}(\boldsymbol{x})A^T\boldsymbol{\lambda}(\boldsymbol{x}). && (15) \end{aligned}$$

Equation (14) implies that the Lagrange multipliers are obtained as the solution of a Gram system

$$G(\boldsymbol{x})\boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}), \tag{16}$$

where the entries of G are the weighted ℓ_2-inner products

$$G_{jk}(\boldsymbol{x}) = \langle p_j, p_k\rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} p_j(\boldsymbol{x}_i)p_k(\boldsymbol{x}_i)W(\boldsymbol{x}, \boldsymbol{x}_i), \quad j, k = 1, \ldots, m. \tag{17}$$

The special feature here is that the weight varies with the evaluation point $\boldsymbol{x}$. The Gram matrix is symmetric and positive definite since the polynomial basis is linearly independent and the weights are positive.

Equation (15) can be written componentwise, i.e., the generating functions in (10) are given by

$$\Psi_i(\boldsymbol{x}) = W(\boldsymbol{x}, \boldsymbol{x}_i)\sum_{j=1}^{m}\lambda_j(\boldsymbol{x})p_j(\boldsymbol{x}_i), \qquad i = 1, \ldots, N. \tag{18}$$

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 (the Gram system) that changes as soon as $\boldsymbol{x}$ changes.

2.2 Standard Interpretation of MLS Approximation

In the standard approach to moving least squares approximation (see [16]) we 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 $\mathcal{U} = \operatorname{span}\{u_1, \ldots, u_m\}$ (with $m < N$), along with the same weighted ℓ_2-inner product

$$\langle f, g \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i), \qquad \boldsymbol{x} \in \mathbb{R}^s \text{ fixed}, \tag{19}$$

as introduced above in (17). Again, the positive weights $W_i = W(\cdot, \boldsymbol{x}_i)$, $i = 1, \ldots, N$, depend on the evaluation point $\boldsymbol{x}$.

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 (19). In the statistics literature this process is known as local regression (see, e.g., the book [8]). Following the usual least squares approach, this means we will obtain the approximation (at the point $\boldsymbol{x}$) as

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x}) u_j(\boldsymbol{x}), \tag{20}$$

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

$$\sum_{i=1}^{N} \left[\mathcal{P}f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i)\right]^2 W_i(\boldsymbol{x}) \tag{21}$$

is minimized. Due to the definition of the inner product (19) whose weight function "moves" with the evaluation point $\boldsymbol{x}$, the coefficients c_j in (20) depend also on $\boldsymbol{x}$. As a consequence one has to solve the normal equations

$$\sum_{j=1}^{m} c_j(\boldsymbol{x}) \langle u_j, u_k \rangle_{W(\boldsymbol{x})} = \langle f, u_k \rangle_{W(\boldsymbol{x})}, \qquad k = 1, \ldots, m, \tag{22}$$

anew each time the evaluation point $\boldsymbol{x}$ is changed. In matrix notation (22) becomes

$$G(\boldsymbol{x}) \boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_u(\boldsymbol{x}), \tag{23}$$

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 (22) all depending on $\boldsymbol{x}$.

Usually one takes $\mathcal{U}$ to be a space of (multivariate) polynomials, i.e.,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} c_j(\boldsymbol{x}) p_j(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^s, \tag{24}$$

where $\{p_1, \ldots, p_m\}$ is a basis for the space Π_d^s of s-variate polynomials of degree d. This restriction is necessary to show equivalence with the Backus-Gilbert approach in the next subsection. However, we will make use of a different space $\mathcal{U}$ when pointing out the connection to RBF approximation later on.

The Gram system (23) now becomes

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x}), \tag{25}$$

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

$$G_{jk}(\boldsymbol{x}) = \langle p_j, p_k \rangle_{W(\boldsymbol{x})} = \sum_{i=1}^{N} p_j(\boldsymbol{x}_i) p_k(\boldsymbol{x}_i) W(\boldsymbol{x}, \boldsymbol{x}_i), \quad j,k = 1, \ldots, m, \tag{26}$$

as in (17), and the right-hand side vector consists of the projections of the data f onto the basis functions, i.e.,

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

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

$$\frac{1}{2} \boldsymbol{c}^T(\boldsymbol{x}) G(\boldsymbol{x}) \boldsymbol{c}(\boldsymbol{x}) - \boldsymbol{\mu}^T(\boldsymbol{x}) \left[G(\boldsymbol{x}) \boldsymbol{c}(\boldsymbol{x}) - A Q^{-1}(\boldsymbol{x}) \boldsymbol{f} \right], \tag{27}$$

where $G(\boldsymbol{x})$ is the Gram matrix (17), $Q(\boldsymbol{x})$ the diagonal matrix of weight functions (13) 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 (25), i.e.,

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

The solution of the linear system resulting from the minimization problem (27) gives us (cf. (5) and (6))

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

so that – analogous to 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 quadratic form

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

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

2.3 Equivalence of the Two Approaches to Moving Least Squares Approximation

We now show that the two approaches to the moving least squares method just described are equivalent, i.e., we show that $\mathcal{P}f(\boldsymbol{x})$ computed via (10) and (24) are the same. On the one hand the standard moving least squares formulation (24) 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 $\boldsymbol{p}(\boldsymbol{x}) = [p_1(\boldsymbol{x}), \ldots, p_m(\boldsymbol{x})]^T$ and $\boldsymbol{c}(\boldsymbol{x}) = [c_1(\boldsymbol{x}), \ldots, c_m(\boldsymbol{x})]^T$.

In (27) we represented 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

$$\mathcal{P}f(\boldsymbol{x}) = \boldsymbol{p}^T(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{p}^T(\boldsymbol{x})G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f}. \tag{28}$$

The approximant (10) in the Backus-Gilbert "ansatz", on the other hand, 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 $\boldsymbol{f} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$. For the Backus-Gilbert approach we derived (see (14) and (15))

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

where $G(\boldsymbol{x}) = AQ^{-1}(\boldsymbol{x})A^T$ (see (17) or (26)). Therefore, we now obtain

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

which, by the symmetry of $Q(\boldsymbol{x})$ and $G(\boldsymbol{x})$, is the same as (28).

2.4 A Dual Representation for the Standard Approach

We just derived that on the one hand (from the Backus-Gilbert formulation)

$$G(\boldsymbol{x})\boldsymbol{\lambda}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}) \quad \Longleftrightarrow \quad \boldsymbol{\lambda}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})\boldsymbol{p}(\boldsymbol{x}). \tag{29}$$

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

$$\Lambda = G^{-1}(\boldsymbol{x})A, \tag{30}$$

where the $m \times N$ matrices Λ and A have entries $\Lambda_{ji} = \lambda_j(\boldsymbol{x}_i)$ and $A_{ji} = p_j(\boldsymbol{x}_i)$, $i = 1, \ldots, N$, $j = 1, \ldots, m$.

The standard formulation, on the other hand, gives us

$$G(\boldsymbol{x})\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{f}_p(\boldsymbol{x}) \quad \Longleftrightarrow \quad \boldsymbol{c}(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})\boldsymbol{f}_p(\boldsymbol{x}) = G^{-1}(\boldsymbol{x})AQ^{-1}(\boldsymbol{x})\boldsymbol{f} \tag{31}$$

where

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

as above. By combining (30) with (31) 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,

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} \langle f, \lambda_j\rangle_{W(\boldsymbol{x})} p_j(\boldsymbol{x}). \tag{32}$$

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

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{m} d_j(\boldsymbol{x})\lambda_j(\boldsymbol{x}) \tag{33}$$

with

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

Moreover, from the linear constraints in the Backus-Gilbert approach (12) we know that the polynomials are reproduced by our projection operator $\mathcal{P}$, i.e.,

$$\mathcal{P}p(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{for all } p \in \Pi_d^s.$$

In the dual approach the Backus-Gilbert minimization requires reproduction of the Lagrange multiplier basis of $\mathcal{U}$, i.e.,

$$\mathcal{P}\lambda_j(\boldsymbol{x}) = \lambda_j(\boldsymbol{x}), \qquad j = 1, \ldots, m.$$

Now we can show that the polynomials p_j and Lagrange multipliers λ_j, $j = 1, \ldots, m$, are *bi-orthonormal*, i.e.,

$$\langle \lambda_k, p_j\rangle_{W(\boldsymbol{x})} = \delta_{jk}, \qquad j, k = 1 \ldots, m. \tag{34}$$

Indeed, if we use λ_k in place of f in the dual expansion (33), then

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

which implies the bi-orthogonality relationship (34) by virtue of the fact that $\mathcal{P}\lambda_k(\boldsymbol{x}) = \lambda_k(\boldsymbol{x})$ established above.

Therefore, the Lagrange multipliers form a basis that is *dual* to the polynomials.

2.5 Summary of MLS Approximation

The equivalence of the Backus-Gilbert approach and the standard approach shows us 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 (even though this is not explicitly enforced by the solution of the normal equations in the standard approach).
- 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 (11) 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) $W_i = W(\cdot, \boldsymbol{x}_i)$.

It is well known that (as a consequence of the Mairhuber-Curtis Theorem) any basis for a nontrivial multivariate Haar space needs to be data dependent. Both the dual Lagrange multiplier basis $\lambda_1, \ldots, \lambda_m$ and the generating functions $\Psi_1, \ldots, \Psi_N$ depend on the data locations $\boldsymbol{x}_i \in \mathcal{X}$ (see also the plots in Section 5).

By also considering the dual expansion (33) we have three alternative representations for the moving least squares quasi-interpolant. This is summarized in the following theorem.

Theorem 2.1 *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 (14). 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 (21) is given by

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

The argumentation used above for polynomials can be generalized for the more general "ansatz" with approximation space $\mathcal{U}$. This 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., [2] or [3]). We will make use of this more general approximation space $\mathcal{U}$ to establish a connection between RBF and MLS approximation in the next section.

3 Alternative Interpretations of Radial Basis Function Approximation

We now return to radial basis functions and show how RBF approximation can be interpreted as a special case of moving least squares approximation. Consequently we will arrive at three alternative formulations for the RBF approximant.

In fact, with the notation introduced thus far this connection is not difficult to see. All we need to do is assume that the approximation space $\mathcal{U}$ in the standard moving least squares approach is given as $\mathcal{U} = \text{span}\{\Phi(\cdot, \boldsymbol{\xi}), \boldsymbol{\xi} \in \Xi\}$, and the weights are identically equal to the constant 1. For simplicity we will restrict our discussion to the case where Ξ is a subset of $\mathcal{X}$. Then the standard MLS formulation with $\mathcal{U}$ specified as above corresponds exactly to the least squares formulation of radial basis function approximation given in the beginning (with the additional identification $m = M$). However, the dual formulation and the Backus-Gilbert approach now provide two additional interpretations of RBF approximation.

3.1 The Backus-Gilbert Approach for RBF Approximation

In the RBF context the Backus-Gilbert approach corresponds to a quasi-interpolant of the form

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{i=1}^{N} f(\boldsymbol{x}_i) K(\boldsymbol{x}, \boldsymbol{x}_i), \tag{35}$$

where as before $\boldsymbol{f} = [f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)]^T$ represents the given data.

Now the values of the generating functions $K(\cdot, \boldsymbol{x}_i)$ are found by minimizing

$$\frac{1}{2} \sum_{i=1}^{N} K^2(\boldsymbol{x}, \boldsymbol{x}_i) \tag{36}$$

subject to the linear constraints

$$\sum_{i=1}^{N} \Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_j) K(\boldsymbol{x}, \boldsymbol{x}_i) = \Phi(\boldsymbol{x}, \boldsymbol{\xi}_j), \quad j = 1, \ldots, M. \tag{37}$$

These constraints imply that we obtain generating functions for the quasi-interpolant (35) that lie in the same space $\mathcal{U}$ that is generated by the usual radial basis functions, i.e., $\mathcal{U} = \text{span}\{\Phi(\cdot, \boldsymbol{\xi}_1), \ldots, \Phi(\cdot, \boldsymbol{\xi}_M)\}$. In fact, any function $u \in \mathcal{U}$ will be reproduced by the projection $\mathcal{P}$, i.e.,

$$\mathcal{P}u(\boldsymbol{x}) = u(\boldsymbol{x}), \quad \text{for all } u \in \mathcal{U}. \tag{38}$$

Moreover, the kernel has the additional optimality property that its ℓ_2-norm is minimized. In particular, for the interpolation problem this implies that the cardinal radial basis functions (see below) have the smallest possible ℓ_2-norm.

The Backus-Gilbert approach also implies that the kernel K is given by

$$K(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{M} \lambda_j(\boldsymbol{x}) \Phi(\boldsymbol{x}, \boldsymbol{y}), \tag{39}$$

where the Lagrange multipliers λ_j, $j = 1, \ldots, M$, form a *dual basis* for $\mathcal{U}$. As before, the dual basis is found by solving the Gram system

$$G\boldsymbol{\lambda} = \boldsymbol{\Phi}, \tag{40}$$

where G has entries

$$G_{jk} = \sum_{i=1}^{N} \Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_j)\Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_k), \quad j, k = 1, \ldots, M,$$

and the vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\Phi}$ collect the dual basis functions and basis functions, respectively.

Note that the dual basis allows us to see the coefficients c_j of the standard RBF approximation

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \Phi(\boldsymbol{x}, \boldsymbol{\xi}_j)$$

(cf. (1)) in a new light. As for the MLS approach we have

$$c_j = \langle f, \lambda_j \rangle, \quad j = 1, \ldots, m.$$

This shows us that the coefficients in the expansion (1) with basis functions $\Phi(\cdot, \boldsymbol{\xi}_j)$ are obtained by projecting the data onto the dual basis. Of course, due to the norm used in the optimization criterion (39) all inner products are now standard ℓ_2-inner products.

The kernel K is symmetric and positive definite. This can easiest be seen by interpreting K as $\boldsymbol{\lambda}^T\boldsymbol{\Phi}$. Then, using the definition (40) of the dual basis we have

$$K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\lambda}^T(\boldsymbol{x})\boldsymbol{\Phi}(\boldsymbol{y}) = \boldsymbol{\lambda}^T(\boldsymbol{x})G\boldsymbol{\lambda}(\boldsymbol{y}),$$

and the symmetry of K follows from the symmetry and positive definiteness of G. Note that we also have the dual interpretation

$$K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\Phi}^T(\boldsymbol{x})G^{-1}\boldsymbol{\Phi}(\boldsymbol{y}).$$

From the Backus-Gilbert approach we know that the radial basis functions are reproduced (see (38)) by our projection operator $\mathcal{P}$, i.e.,

$$\mathcal{P}\Phi(\boldsymbol{x}, \boldsymbol{\xi}_j) = \Phi(\boldsymbol{x}, \boldsymbol{\xi}_j), \qquad j = 1, \ldots, M.$$

A Backus-Gilbert minimization can also be formulated for the dual basis. This implies the reproduction

$$\mathcal{P}\lambda_j(\boldsymbol{x}) = \lambda_j(\boldsymbol{x}), \qquad j = 1, \ldots, M,$$

and, as before, we have bi-orthonormality of the basis functions $\Phi_j = \Phi(\cdot, \boldsymbol{\xi}_j)$ and dual basis functions λ_j, $j = 1, \ldots, M$, i.e.,

$$\langle \lambda_k, \Phi_j \rangle = \delta_{jk}, \qquad j, k = 1 \ldots, M. \tag{41}$$

We claimed above that in the special case $M = N$ with $\Xi = \mathcal{X}$ the quasi-interpolant becomes an interpolant, i.e., the functions $K(\cdot, \boldsymbol{x}_i)$ form a *cardinal basis* for $\mathcal{U}$. This

claim is equivalent to showing that for $M = N$ we have $K(\boldsymbol{x}_i, \boldsymbol{x}_\ell) = \delta_{i\ell}$. To see this, we first note that the radial basis functions are symmetric, i.e., $\Phi(\boldsymbol{x}, \boldsymbol{\xi}_j) = \Phi(\boldsymbol{\xi}_j, \boldsymbol{x})$. Since the dual basis functions are obtained as linear combinations of the basis functions they are also symmetric. Therefore,

$$\begin{aligned} K(\boldsymbol{x}_i, \boldsymbol{x}_\ell) &= \sum_{j=1}^{n} \lambda_j(\boldsymbol{x}_i)\Phi_j(\boldsymbol{x}_\ell) \\ &= \sum_{j=1}^{n} \lambda_i(\boldsymbol{\xi}_j)\Phi_\ell(\boldsymbol{\xi}_j) \\ &= \langle \lambda_i, \Phi_\ell \rangle = \delta_{i\ell} \end{aligned}$$

where the last step requires equality of Ξ with $\mathcal{X}$ and the bi-orthogonality relation (41).

Moreover,

$$\begin{aligned} \mathcal{P}f(\boldsymbol{x}) &= \sum_{i=1}^{N} f(\boldsymbol{x}_i)K(\boldsymbol{x}, \boldsymbol{x}_i) \\ &= \sum_{i=1}^{N} f(\boldsymbol{x}_i) \sum_{j=1}^{M} \lambda_j(\boldsymbol{x})\Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_j) \\ &= \sum_{j=1}^{M} \underbrace{\left[\sum_{i=1}^{N} f(\boldsymbol{x}_i)\Phi(\boldsymbol{x}_i, \boldsymbol{\xi}_j) \right]}_{=\langle f, \Phi_j \rangle = d_j} \lambda_j(\boldsymbol{x}). \end{aligned}$$

Thus, we have the dual expansion

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{M} d_j \lambda_j(\boldsymbol{x}),$$

where the coefficients d_j are given as projections of the data onto the original basis functions $\Phi_j = \Phi(\cdot, \boldsymbol{\xi}_j)$.

The following theorem summarizes our different RBF representations.

Theorem 3.1 *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 $\Xi = \{\boldsymbol{\xi}_j\}_{j=1}^M \subset \mathbb{R}^s$ be another set of points that generates the approximation space $\mathcal{U} = \operatorname{span}\{\Phi(\cdot, \boldsymbol{\xi}_j),\ \boldsymbol{\xi}_j \in \Xi\}$. Furthermore, let λ_j, $j = 1, \ldots, M$, be the dual basis functions defined by (40) and K the kernel given by (39). The best least squares approximation to f on $\mathcal{X}$ from $\mathcal{U}$ with respect to the standard ℓ_2-inner product is given by*

$$\mathcal{P}f(\boldsymbol{x}) = \sum_{j=1}^{M} \langle f, \lambda_j \rangle \Phi_j(\boldsymbol{x}) \tag{42}$$

$$= \sum_{j=1}^{M} \langle f, \Phi_j \rangle \lambda_j(\boldsymbol{x}) \tag{43}$$

$$= \sum_{i=1}^{N} f(\boldsymbol{x}_i)K(\boldsymbol{x}, \boldsymbol{x}_i). \tag{44}$$

Since the space $\mathcal{U}$ is only M-dimensional, Theorem 3.1 implies that, for $M < N$, the set $\{K(\cdot, \boldsymbol{x}_i)\}_{i=1}^N$ is linearly dependent on Ξ. In other words, the sets $\{\Phi_j\}_{j=1}^M$ and $\{\lambda_j\}_{j=1}^N$ are bases for $\mathcal{U}$, whereas $\{K(\cdot, \boldsymbol{x}_i)\}_{i=1}^N$ is only a frame, i.e., contains redundancies.

4 Reproducing Kernels

In the introduction we mentioned the reproducing kernel Hilbert space (native space) associated with the strictly positive definite kernel Φ. We now want to study the reproducing properties of the kernels K and Ψ introduced above. We start with the precise definition of a reproducing kernel (see, e.g., [6] or [30]).

Definition 4.1 *Let Ω be a (rather arbitrary) domain, and $\mathcal{H}$ be a real Hilbert space of functions $f : \Omega \to \mathbb{R}$. A function $K : \Omega \times \Omega \to \mathbb{R}$ is called* reproducing kernel for $\mathcal{H}$ *if*

1. *$K(\cdot, \boldsymbol{x}) \in \mathcal{H}$ for all $\boldsymbol{x} \in \Omega$,*
2. *$f(\boldsymbol{x}) = \langle f, K(\cdot, \boldsymbol{x})\rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$ and all $\boldsymbol{x} \in \Omega$.*

It is also known that the reproducing kernel K is strictly positive definite and unique.

4.1 Discrete RBF Reproducing Kernels

In the RBF setting we consider $\Omega = \mathcal{X}$ and the space $\mathcal{U} = \operatorname{span}\{\Phi(\cdot, \boldsymbol{\xi}_1), \ldots, \Phi(\cdot, \boldsymbol{\xi}_M)\}$. The Backus-Gilbert formulation ensures that the projection $\mathcal{P}$ reproduces every function $u \in \mathcal{U}$ (see (38)). Moreover, since $K(\cdot, \boldsymbol{x}) \in \mathcal{U}$ for any $\boldsymbol{x} \in \mathcal{X}$ (by virtue of (39)) and

$$\langle u, K(\cdot, \boldsymbol{x})\rangle = \sum_{i=1}^{N} u(\boldsymbol{x}_i) K(\boldsymbol{x}_i, \boldsymbol{x}) = \mathcal{P}u(\boldsymbol{x}) = u(\boldsymbol{x})$$

for all $u \in \mathcal{U}$ we see that K is the reproducing kernel of $\mathcal{U}$ with respect to the discrete ℓ_2-norm. Therefore, we refer to K as a *discrete reproducing kernel*. It should be pointed out that the space $\mathcal{U}$ is a space that is much different from the native space $\mathcal{N}_\Phi$ of Φ. It is defined on a discrete domain, and with respect to a discrete inner product.

We can also consider an equivalent inner product for functions in $\mathcal{U}$. To this end, we express the first function as $f = \sum_{j=1}^M c_j \Phi_j$, and the second function in the dual representation $g = \sum_{j=1}^M d_j \lambda_j$. Then we define the inner product

$$\langle f, g\rangle_{\mathcal{U}} = \sum_{j=1}^{M} c_j d_j. \tag{45}$$

In the special case of interpolation (i.e., $M = N$ with $\Xi = \mathcal{X}$) this inner product is equivalent to the discrete ℓ_2-inner product, i.e.,

$$\langle f, g\rangle_{\mathcal{U}} = \langle f, g\rangle.$$

Since any function in $\mathcal{U}$ is reproduced by $\mathcal{P}$ we know that $c_j = \langle f, \lambda_j \rangle$ and $d_j = \langle g, \Phi_j \rangle$. Thus, using the definition of the ℓ_2-inner product and (39) we have

$$\begin{aligned}
\sum_{j=1}^{M} c_j d_j &= \sum_{j=1}^{M} \langle f, \lambda_j \rangle \langle g, \Phi_j \rangle, \\
&= \sum_{j=1}^{M} \sum_{i=1}^{N} f(\boldsymbol{x}_i) \lambda_j(\boldsymbol{x}_i) \sum_{\ell=1}^{N} g(\boldsymbol{x}_\ell) \Phi_j(\boldsymbol{x}_\ell) \\
&= \sum_{i=1}^{N} \sum_{\ell=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_\ell) \sum_{j=1}^{M} \lambda_j(\boldsymbol{x}_i) \Phi_j(\boldsymbol{x}_\ell) \\
&= \sum_{i=1}^{N} \sum_{\ell=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_\ell) K(\boldsymbol{x}_i, \boldsymbol{x}_\ell).
\end{aligned}$$

Now, making use of the cardinality property $K(\boldsymbol{x}_i, \boldsymbol{x}_\ell) = \delta_{i\ell}$ for the interpolation case, we obtain

$$\begin{aligned}
\sum_{i=1}^{N} \sum_{\ell=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_\ell) K(\boldsymbol{x}_i, \boldsymbol{x}_\ell) &= \sum_{i=1}^{N} \sum_{\ell=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_\ell) \delta_{i\ell} \\
&= \sum_{i=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_i) \\
&= \langle f, g \rangle.
\end{aligned}$$

4.2 Discrete MLS Reproducing Kernels

In the MLS setting the kernel Ψ was given by (see (18))

$$\Psi(\boldsymbol{x}, \boldsymbol{y}) = W(\boldsymbol{x}, \boldsymbol{y}) \sum_{j=1}^{m} \lambda_j(\boldsymbol{x}) p_j(\boldsymbol{y}), \qquad \boldsymbol{y} \in \mathcal{X}.$$

Due to the presence of the weight function the definition of an inner product and verification of reproduction properties is complicated. We again present two different approaches. First, we directly consider functions in the space spanned by Ψ on $\mathcal{X}$. Our second approach is to decouple the weight function using Fourier series. This leads to a discrete framework that is *infinite-dimensional*.

For functions of the form

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i \Psi(\boldsymbol{x}, \boldsymbol{x}_i), \qquad g(\boldsymbol{x}) = \sum_{\ell=1}^{N} \beta_\ell \Psi(\boldsymbol{x}, \boldsymbol{x}_\ell),$$

with $\boldsymbol{x}_i \in \mathcal{X}$ and arbitrary (real) coefficients α_i and β_ℓ we can define the inner product

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{N} \sum_{\ell=1}^{N} \alpha_i \beta_\ell \Psi(\boldsymbol{x}_i, \boldsymbol{x}_\ell). \tag{46}$$

This definition implies

$$\langle \Psi(\boldsymbol{x}, \cdot), \Psi(\cdot, \boldsymbol{y}) \rangle_{\mathcal{H}} = \Psi(\boldsymbol{x}, \boldsymbol{y}), \qquad \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}.$$

The norm in $\mathcal{H}$ is then given by

$$\|f\|_{\mathcal{H}}^2 = \langle f, f\rangle_{\mathcal{H}} = \sum_{i=1}^{N}\sum_{\ell=1}^{N} \alpha_i \alpha_\ell \Psi(\boldsymbol{x}_i, \boldsymbol{x}_\ell).$$

It is clear that this quantity is positive (for nonzero f) since Ψ is symmetric and positive definite.

The reproducing property of $\Psi(\cdot, \boldsymbol{x})$ for functions in $\mathcal{H}$ follows from

$$\begin{aligned}
\langle f, \Psi(\cdot, \boldsymbol{x})\rangle_{\mathcal{H}} &= \langle \sum_{i=1}^{N} \alpha_i \Psi(\boldsymbol{x}_i, \cdot), \Psi(\cdot, \boldsymbol{x})\rangle_{\mathcal{H}} \\
&= \sum_{i=1}^{N} \alpha_i \Psi(\boldsymbol{x}_i, \boldsymbol{x}) \\
&= f(\boldsymbol{x}).
\end{aligned}$$

Also, clearly, the function $\Psi(\cdot, \boldsymbol{x})$ is an element of $\mathcal{H}$ whenever $\boldsymbol{x} \in \mathcal{X}$.

For our second interpretation we now assume that the weight function is not just positive, but positive definite and translation invariant so that we can apply Bochner's Theorem. First we represent the weight function by its (multivariate) Fourier series, i.e.,

$$\begin{aligned}
W(\boldsymbol{x}, \boldsymbol{y}) = w(\boldsymbol{x} - \boldsymbol{y}) &= \sum_{\boldsymbol{k}\in\mathbb{Z}^s} \mu_{\boldsymbol{k}} e^{2\pi i \boldsymbol{k}(\boldsymbol{x}-\boldsymbol{y})} \\
&= \sum_{\boldsymbol{k}\in\mathbb{Z}^s} \mu_{\boldsymbol{k}} e^{2\pi i \boldsymbol{k}\boldsymbol{x}} e^{-2\pi i \boldsymbol{k}\boldsymbol{y}} \\
&= \sum_{\boldsymbol{k}\in\mathbb{Z}^s} \mu_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}(\boldsymbol{x}) \overline{\varphi}_{\boldsymbol{k}}(\boldsymbol{y})
\end{aligned}$$

with

$$\mu_{\boldsymbol{k}} = \hat{w}(\boldsymbol{k}) = \int_{\mathbb{R}^s} w(\boldsymbol{x}) e^{-2\pi i \boldsymbol{k}\boldsymbol{x}} d\boldsymbol{x}.$$

Note that Bochner's Theorem ensures the nonnegativity of the coefficients $\mu_{\boldsymbol{k}}$, and therefore absolute convergence of the Fourier series. Moreover, the difference between various weight functions is given only by these coefficients. The Fourier terms are the same for all weight functions.

Now we can use this (absolutely convergent) Fourier series in the representation of the reproducing kernel. This gives us

$$\begin{aligned}
\Psi(\boldsymbol{x}, \boldsymbol{y}) &= W(\boldsymbol{x}, \boldsymbol{y}) \sum_{j=1}^{m} \lambda_j(\boldsymbol{x}) p_j(\boldsymbol{y}) \\
&= \sum_{\boldsymbol{k}\in\mathbb{Z}^s} \mu_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}(\boldsymbol{x}) \overline{\varphi}_{\boldsymbol{k}}(\boldsymbol{y}) \sum_{j=1}^{m} \lambda_j(\boldsymbol{x}) p_j(\boldsymbol{y}) \\
&= \sum_{k=1}^{\infty} \nu_k \rho_k(\boldsymbol{x}) \sigma_k(\boldsymbol{y}).
\end{aligned}$$

After this re-arrangement (and change from counting on the infinite lattice to indexing over positive integers), m different coefficients ν_k are equal to one $\mu_{\boldsymbol{k}}$, and the functions ρ_k are products of $\varphi_{\boldsymbol{k}}$ and λ_j, whereas σ_k is built from $\overline{\varphi}_{\boldsymbol{k}}$ and p_j.

We can now observe that for any $f \in \mathcal{H}$ (the same $\mathcal{H}$ as in our first approach)

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{i=1}^{N} \alpha_i \Psi(\boldsymbol{x}, \boldsymbol{x}_i) \\ &= \sum_{i=1}^{N} \alpha_i \Psi(\boldsymbol{x}_i, \boldsymbol{x}) \\ &= \sum_{i=1}^{N} \alpha_i \sum_{k=1}^{\infty} \nu_k \rho_k(\boldsymbol{x}_i) \sigma_k(\boldsymbol{x}) \\ &= \sum_{k=1}^{\infty} \nu_k \left[\sum_{i=1}^{N} \alpha_i \rho_k(\boldsymbol{x}_i) \right] \sigma_k(\boldsymbol{x}) \\ &= \sum_{k=1}^{\infty} \nu_k \langle \alpha, \rho_k \rangle \sigma_k(\boldsymbol{x}) = \sum_{k=1}^{\infty} \gamma_k \sigma_k(\boldsymbol{x}). \end{aligned}$$

Here, α represents the function whose values on the set $\mathcal{X}$ correspond to the coefficients α_i, $i = 1, \ldots, N$. Moreover, we have introduced the abbreviation

$$\gamma_k = \nu_k \langle \alpha, \rho_k \rangle, \quad k = 1, 2, \ldots,$$

where

$$\langle f, g \rangle = \sum_{i=1}^{N} f(\boldsymbol{x}_i) g(\boldsymbol{x}_i)$$

is the (uniformly weighted) discrete ℓ_2-inner product on $\mathcal{X}$ as used in the RBF section.

We also have a dual interpretation (expressed in terms of a function g instead of f)

$$\begin{aligned} g(\boldsymbol{x}) &= \sum_{\ell=1}^{N} \beta_\ell \Psi(\boldsymbol{x}, \boldsymbol{x}_\ell) \\ &= \sum_{\ell=1}^{N} \beta_\ell \sum_{k=1}^{\infty} \nu_k \rho_k(\boldsymbol{x}) \sigma_k(\boldsymbol{x}_\ell) \\ &= \sum_{k=1}^{\infty} \nu_k \left[\sum_{\ell=1}^{N} \beta_\ell \sigma_k(\boldsymbol{x}_\ell) \right] \rho_k(\boldsymbol{x}) \\ &= \sum_{k=1}^{\infty} \nu_k \langle \beta, \sigma_k \rangle \rho_k(\boldsymbol{x}) = \sum_{k=1}^{\infty} \delta_k \rho_k(\boldsymbol{x}). \end{aligned}$$

Here

$$\delta_k = \nu_k \langle \beta, \sigma_k \rangle, \quad k = 1, 2, \ldots,$$

and β plays the same role as α above.

Note that now we have a finite series representation for functions in $\mathcal{H}$ (with basis $\Psi(\cdot, \boldsymbol{x}_i)$, $i = 1, \ldots, N$) and two infinite series representations (with *frame* ρ_k, or *dual frame* σ_k, $k = 1, 2, \ldots$).

Again, there are three equivalent representations of the least squares approximation.

Theorem 4.2 *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 the approximation space be generated by the same set, i.e., $\mathcal{H} = \text{span}\{\Psi(\cdot, \boldsymbol{x}_i),\ \boldsymbol{x}_i \in \mathcal{X}\}$, where Ψ is the kernel given by (18). Furthermore, let ρ_k and σ_k, $k = 1, 2, \ldots$, be the frame and dual frame defined above. The best least squares approximation to f on $\mathcal{X}$ from $\mathcal{H}$ with respect to the inner product (19) is given by*

$$\begin{aligned} \mathcal{P}f(\boldsymbol{x}) &= \sum_{i=1}^{N} f(\boldsymbol{x}_i)\Psi(\boldsymbol{x}, \boldsymbol{x}_i) && (47) \\ &= \sum_{k=1}^{\infty} \nu_k \langle f, \sigma_k \rangle \rho_k(\boldsymbol{x}) && (48) \\ &= \sum_{k=1}^{\infty} \nu_k \langle f, \rho_k \rangle \sigma_k(\boldsymbol{x}). && (49) \end{aligned}$$

The Fourier series representation also provides us with an equivalent formulation for the Hilbert space inner product (46) defined above, i.e.,

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{k=1}^{\infty} \frac{\gamma_k \delta_k}{\nu_k}. \tag{50}$$

To see this we begin with the expression given on the right-hand side of the formula and derive (46):

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{\gamma_k \delta_k}{\nu_k} &= \sum_{k=1}^{\infty} \frac{\nu_k \langle \alpha, \rho_k \rangle \nu_k \langle \beta, \sigma_k \rangle}{\nu_k} \\ &= \sum_{k=1}^{\infty} \nu_k \langle \alpha, \rho_k \rangle \langle \beta, \sigma_k \rangle \\ &= \sum_{k=1}^{\infty} \nu_k \sum_{i=1}^{N} \alpha_i \rho_k(\boldsymbol{x}_i) \sum_{\ell=1}^{N} \beta_\ell \sigma_k(\boldsymbol{x}_\ell) \\ &= \sum_{i=1}^{N} \sum_{\ell=1}^{N} \alpha_i \beta_\ell \sum_{k=1}^{\infty} \nu_k \rho_k(\boldsymbol{x}_i) \sigma_k(\boldsymbol{x}_\ell) \\ &= \sum_{i=1}^{N} \sum_{\ell=1}^{N} \alpha_i \beta_\ell \Psi(\boldsymbol{x}_i, \boldsymbol{x}_\ell) \\ &= \langle f, g \rangle_{\mathcal{H}}. \end{aligned}$$

Note that the formula given in (50) is the analogue of the finite formula (45) we had in the RBF setting.

5 Examples

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 [24]. 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 (cf. (26))

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

implies

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

The dual basis is given by (see (16))

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

so that

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

and

$$\mathcal{P}f(\boldsymbol{x}) = d_1(\boldsymbol{x})\lambda_1(\boldsymbol{x}) \tag{51}$$

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 reproducing kernel is defined as (see (18))

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

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

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

Of course this is the same as the basis expansion $c_1(\boldsymbol{x})$ and the dual expansion (51).

We should now have bi-orthonormality on $\mathcal{X}$, i.e.,

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

Indeed,

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

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

5.2 Plots of Basis-Dual Basis Pairs

To obtain plots of a typical Gaussian basis function and dual basis function for the RBF framework we let $\Xi = \mathcal{X}$ be the set of 13 equally spaced points in $[-5, 5]$.

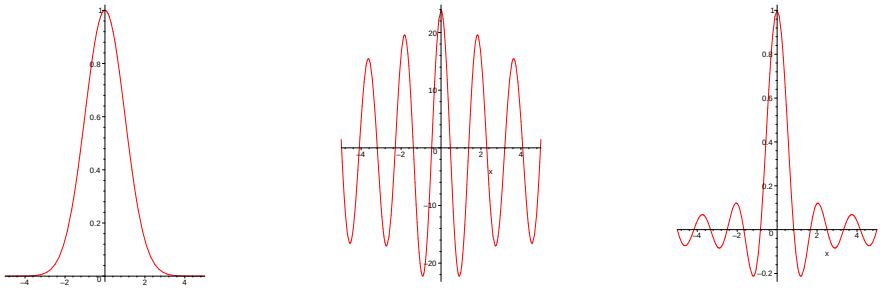

Figure 1: Plot of Gåussian basis function (left), dual basis (center), and reproducing kernel (right) centered at $\boldsymbol{\xi}_7 = 0$.

Figure 1 shows these functions along with a corresponding discrete reproducing kernel K associated with the center $\boldsymbol{\xi}_7 = 0$ in the middle of the interval, whereas Figure 2 shows the functions centered at the left endpoint $\boldsymbol{\xi}_1 = -5$.

Notice the scale on the dual basis as well as the fact that the kernel is a cardinal function, i.e., equal to 1 at $\boldsymbol{\xi}_7 = 0$, and equal to zero at all other $\boldsymbol{\xi}_j$. In the case of true approximation, i.e., when $M < N$, the plots will look similar. However, the kernel is no longer a cardinal function – but only an approximate cardinal function (a generating function for a quasi-interpolant).

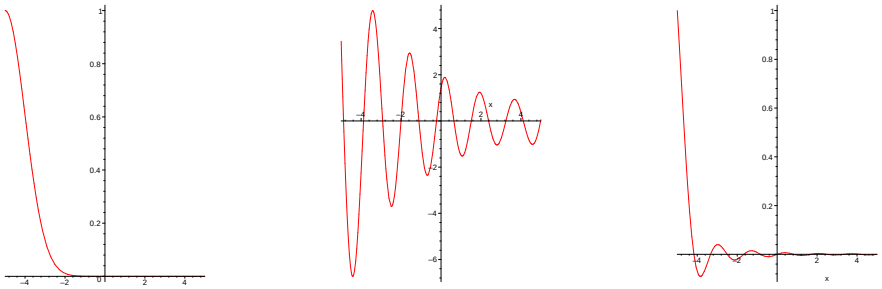

Figure 2: Plot of Gåussian basis function (left), dual basis (center), and reproducing kernel (right) centered at $\boldsymbol{\xi}_1 = -5$.

A set of basis, dual basis and reproducing kernel for interpolation at 13 non-uniformly spaced points in $[-5, 5]$ is shown in Figure 3. The dependence of the dual basis and reproducing kernel on the data is clearly visible.

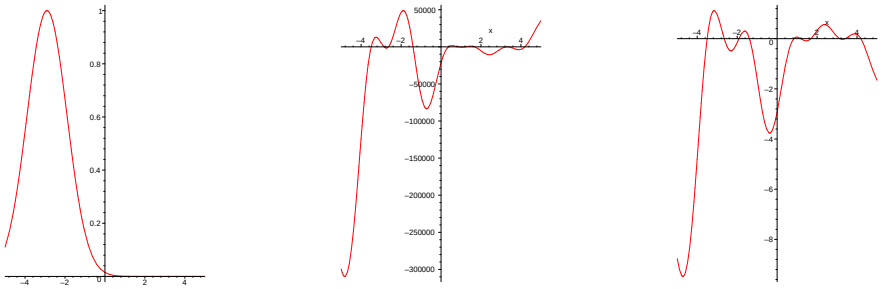

Figure 3: Plot of Gaussian basis function (left), dual basis (center), and reproducing kernel (right) centered at $\boldsymbol{\xi}_1 = -5$.

For the plots in the moving least squares case we again let $\mathcal{X}$ be the set of 13 equally spaced points in $[-5, 5]$. However, now $m = 2$, i.e., we consider the case that ensures reproduction of quadratic polynomials. The weight function is taken to be the same Gaussian as above.

The three basis polynomials $p_1(x) = 1$, $p_2(x) = x$, and $p_3(x) = x^2$ are shown in Figure 4, whereas the dual basis functions λ_1, λ_2, and λ_3 are displayed in Figure 5. The figure shows that, except for the boundary effects caused by the finite interval, these functions resemble a quadratic, linear and constant polynomial.

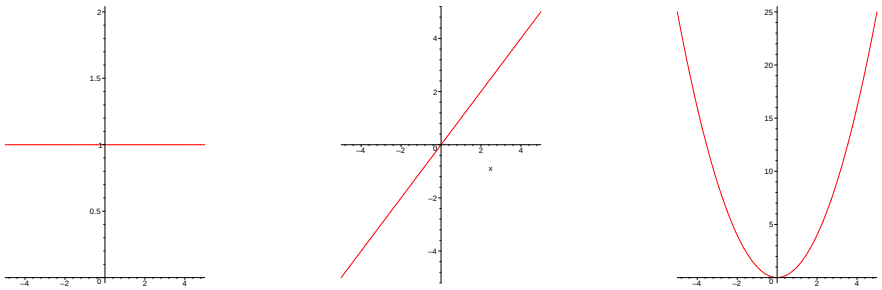

Figure 4: Plot of three polynomial basis functions for MLS approximation.

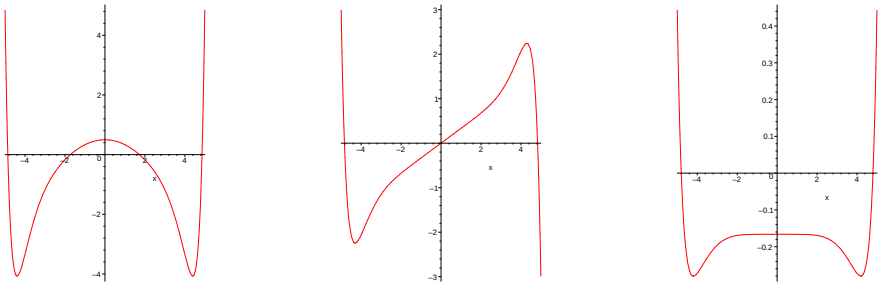

Figure 5: Plot of three dual basis functions for MLS approximation.

In Figure 6 we plot one of the reproducing kernels (centered at $\boldsymbol{\xi}_7 = 0$) along with

an approximate MLS 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 [11].

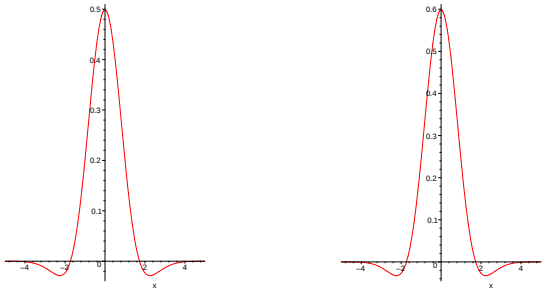

Figure 6: Plot of MLS kernel (left) and approximate MLS generating function (right) centered at $\boldsymbol{x}_7 = 0$.

6 Summary

In this paper we attempted to provide a unified framework for both RBF and MLS approximation. This framework turned out to be the setting of a linearly constrained quadratic minimization problem. The two approaches to MLS approximation present in the literature (Backus-Gilbert optimization, and weighted moving least squares approximation) were transferred to the setting of RBF approximation. This showed that RBF approximation can be considered as a special case of MLS approximation (with global uniform weights), and also resulted in new interpretations of RBF approximation. Certain discrete reproducing kernel spaces were also discussed.

References

[1] S. N. Atluri and S. Shen, *The Meshless Local Petrov-Galerkin (MLPG) Method*, Tech Science Press, Encino, CA, 2002.

[2] I. Babuška and M. Melenk, *The partition of unity method*, Int. J. Numer. Meths. Eng. (40) 1997, 727–758.

[3] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming and P. Krysl, *Meshless methods: an overview and recent developments*, Comp. Meth. Appl. Mech. Eng. (139) 1996, 3–47.

[4] L. P. Bos and K. Šalkauskas, Moving least-squares are Backus-Gilbert optimal, J. Approx. Theory 59 (1989), 267–275.

[5] M. D. Buhmann, *Radial Basis Functions : Theory and Implementations*, Cambridge University Press, Cambridge, 2003.

[6] E. W. Cheney and W. A. Light, *A Course in Approximation Theory*, Brooks/Cole, Pacific Grove, CA, 1999.

[7] T. Evgeniou, M. Pontil, and T. Poggio, Regularization networks and support vector machines, Adv. Comput. Math. 13 (2000), no. 1, 1–50.

[8] J. Fan and I. Gijbels, *Local Polynomial Modelling and its Applications*, Chapman & Hall (New York), 1996.

[9] G. E. Fasshauer, Matrix-free multilevel moving least-squares methods, in *Approximation Theory, X* (St. Louis, MO, 2001), C. K. Chui, L. L. Schumaker, and J. Stöckler (eds.), Innov. Appl. Math., Vanderbilt Univ. Press, Nashville, TN, 2002, 271–281.

[10] G. E. Fasshauer, Approximate moving least-squares approximation with compactly supported radial weights, in *Lecture Notes in Computer Science and Engineering, Vol. 26: Meshfree Methods for Partial Differential Equations*, M. Griebel and M. A. Schweitzer (eds.), Springer, Berlin, 2002, 105–116.

[11] G. E. Fasshauer, Approximate moving least-squares approximation: a fast and accurate multivariate approximation method, in *Curve and Surface Fitting: Saint-Malo 2002*, A. Cohen, J.-L. Merrien, and L. L. Schumaker (eds.), Nashboro Press, 2003, 139–148.

[12] F. Girosi, An equivalence between sparse approximation and support vector machines, Neural Computation 10 (1998), 1455–1480.

[13] M. von Golitschek and L. L. Schumaker, Data fitting by penalized least squares, in Algorithms for Approximation II, M. G. Cax and J. C. Mason (eds.), Chapman & Hall, London, 1990, 210–227.

[14] J. Kybic, T. Blu, and M. Unser, Generalized sampling: A variational approach – Part I: Theory, IEEE Trans. Signal Proc., 50 (2002), 1965–1976.

[15] J. Kybic, T. Blu, and M. Unser, Generalized sampling: A variational approach – Part II: Applications, IEEE Trans. Signal Proc., 50 (2002), 1977–1985.

[16] P. Lancaster and K. Šalkauskas, *Surfaces generated by moving least squares methods*, Math. Comp. (37) 1981, 141–158.

[17] G. R. Liu, *Mesh Free Methods: Moving beyond the Finite Element Method*, CRC Press, Boca Raton, FL, 2002.

[18] C. H. Reinsch, Smoothing by spline functions, Numer. Math. 10 (1967), 177–183.

[19] R. Schaback, *Optimal recovery in translation-invariant spaces of functions*, Annals of Numerical Mathematics, 4 (1997), 547–556.

[20] R. Schaback, Native Hilbert spaces for radial basis functions. I, in *New Developments in Approximation Theory* (Dortmund, 1998), Internat. Ser. Numer. Math., 132, Birkhäuser, Basel, 1999, 255–282.

[21] R. Schaback, A unified theory of radial basis functions. Native Hilbert spaces for radial basis functions. II, Numerical analysis in the 20th century, Vol. I, Approximation theory, J. Comput. Appl. Math. 121 (2000), no. 1-2, 165–177.

[22] R. Schaback and J. Werner, Linearly constrained reconstruction of functions by kernels with applications to machine learning, preprint, Universität Göttingen, 2003.

[23] I. J. Schoenberg, Spline functions and the problem of graduation, Proc. Nat. Acad. Sci. 52 (1964), 947–950.

[24] D. Shepard, A two dimensional interpolation function for irregularly spaced data, Proc. 23rd Nat. Conf. ACM (1968), 517–523.

[25] N. Sivakumar and J. D. Ward, On the least squares fit by radial functions to multidimensional scattered data, Numer. Math. 65 (1993), 219–243.

[26] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997.

[27] M. Unser, Sampling – 50 years after Shannon, Proc. IEEE 88 (2000), 569–587.

[28] G. Wahba, Convergence rate of "thin plate" smoothing splines when the data are noisy (preliminary report), Springer Lecture Notes in Math., 757, 1979, 233–245.

[29] G. Wahba and Z. Luo, Smoothing spline ANOVA fits for very large, nearly regular data sets, with application to historical global climate data, in *The Heritage of P. L. Chebyshev: a Festschrift in honor of the 70th birthday of T. J. Rivlin*, Ann. Numer. Math. 4 (1997), no. 1-4, 579–597.

[30] H. Wendland, *Scattered Data Modelling by Radial and Related Functions*, Habilitation thesis, Univ. Göttingen, 2002.

[31] E. T. Whittaker, On a new method of graduation, Proc. Edinburgh Math. Soc. 41 (1923), 63–75.