Special Techniques for Kernel-Based Reconstruction of Functions from Meshless Data

R. Schaback[(1)]

Abstract: *Here are three short stories on meshless methods using kernel techniques:*

- *Any well–posed linear problem in the native space $\mathcal{N}_\Phi$ of a symmetric (strictly) positive definite kernel Φ can be successfully solved by symmetric meshless collocation. This applies to a large variety of standard linear PDE problems.*
- *Relaxing interpolation conditions by allowing some small absolute error can significantly reduce the complexity of meshless techniques, in particular in conjunction with greedy methods and learning algorithms.*
- *The instability phenomena of badly scaled meshless techniques for smooth kernels can be overcome by an unexpected link to multivariate polynomial interpolation. In particular, there is a preconditioning technique that completely removes the instability in the limit and has a surprisingly simple form, separating the scale information from the geometric information.*

Since the readers can be assumed to be familiar with basic notions of meshless methods, and since detailed presentations are given in the references, it suffices to give a commented overview and suggestions for future research.

1 Short Story on Meshless Kernel Collocation

We assert here that all reasonable analytic problems can in principle be solved numerically by meshless symmetric collocation using smooth positive definite kernels.
Assume that a user has to find a function u that solves a very general analytic problem of the form

$$L_i(u) = f_i \text{ on some set } \Omega_i \subset \mathbb{R}^d,\ 1 \le i \le K \tag{1}$$

where the linear operators L_i may be of any type. Note that the Poisson problem and many, many others take this form for a mixed choice of operators within (parts of) domains and (parts of) boundaries. Assume further that there is a (strictly) positive definite [11, 13, 10]

[1]University of Göttingen, Germany (schaback@math.uni-goettingen.de).

kernel $\Phi(x,y)$ on $\mathbb{R}^d$ which is smooth enough such that all kernels $\Phi_{L_i}(x,y) := L_i^x L_i^y \Phi(x,y)$ are continuous on Ω_i and such that all linear functionals $\delta_x L_i$ are in the dual $\mathcal{N}_\Phi^*$ of the native [4, 5] space $\mathcal{N}_\Phi$ of Φ. In notations like $L_i^x L_i^y \Phi(x,y)$ we use the upper index to denote the variable on which the operator acts. Observe that the Gaussian $\Phi(x,y) = \exp(-\|x-y\|_2^2)$ is a radial basis function that satisfies these assumptions for almost all reasonable linear operators.

If we put all data of (1) simultaneously into a single problem, we have to reconstruct a function u from its given data

$$\lambda(u) \in \mathbb{R},\ \lambda \in \Lambda \subseteq \mathcal{N}_\Phi^*. \tag{2}$$

where λ varies over an uncountably infinite set $\Lambda \subset \mathcal{N}_\Phi^*$ of linear functionals. The symmetric collocation technique of Z.M. Wu [14] then approximates u by linear combinations of functions $\lambda^x \Phi(x,\cdot)$ where λ varies over a finite subset of Λ. Note that we distinguish this technique from the unsymmetric collocation method of Ed Kansa [3].

Theorem. [8] *If there is some function $u \in \mathcal{N}_\Phi$ that satisfies (1), and if symmetric collocation is performed on finite and asymptotically dense subsets of the Ω_i, then the collocation solutions converge in $\mathcal{N}_\Phi$ to a function $\tilde{u} \in \mathcal{N}_\Phi$ which satisfies (1). If the analytic problem (1) is uniquely solvable (i.e. if the homogeneous problem has only the trivial solution), then symmetric collocation recovers u in the limit.*

The proof technique of [8] uses the minimal norm interpolation property to handle countable sets Λ first. The transition to uncountably many data via density is then reduced to the continuity of kernels $L_i^x L_i^y \Phi(x,y)$.

Summarizing, any well–posed linear problem in the native space $\mathcal{N}_\Phi$ of a symmetric (strictly) positive definite kernel Φ can be successfully solved by symmetric meshless collocation. However, it may be a serious drawback that the data must be provided by some unknown function u from $\mathcal{N}_\Phi$. Indeed, many of the successful applications of meshelss kernel methods use data coming from very smooth functions. Since the native spaces of smooth kernels are very small, the above assumption may ask for far too much regularity of the problem.

But in many cases one can replace the original data $\lambda(u)$ of a nonsmooth function u by data coming from a smooth close approximation of u. This replaces the given problem by a slightly perturbed problem that satisfies the assumptions of this section. In other words, meshless symmetric collocation will in practice solve a nearby problem with smooth data. But it will require some future research to make the above statement precise.

2 Short Story on Reduction Techniques

To continue the above story in another direction, let us again assume that we have to recover a function u from the native space $\mathcal{N}_\Phi$ of some (strictly) positive definite kernel $\Phi(x,y)$ from possibly infinite given data (2). In principle, any approximate solution obtained by symmetric collocation will be a linear combination of very many functions of the form $\lambda^x \Phi(x,\cdot)$ for $\lambda \in \Lambda$. But this is not desirable, if the number of required functionals is large and if the result has to be evaluated in very many points.

Borrowing an idea from the theory of learning algorithms and support vector machines [12] and relying on the minimal–norm property of generalized interpolation in native spaces, we fix some $\varepsilon > 0$ and consider the relaxed problem

$$\min\{\|v\|_\Phi \;:\; |\lambda(u-v)| \le \varepsilon \text{ for all } \lambda \in \Lambda\}.$$

It takes the form of a linearly constrained quadratic optimization problem

$$\begin{array}{c} \min \|v\|_\Phi^2 \text{ for all } v \in \mathcal{N}_\Phi \text{ with} \\ \lambda(u) - \varepsilon \le \lambda(v) \le \lambda(u) + \varepsilon \text{ for all } \lambda \in \Lambda. \end{array} \tag{3}$$

By Kuhn–Tucker theory, the solution is based on a subset of "active" constraints. In this context, the solution of the relaxed problem is in the span of a subset of functions $\lambda^x\Phi(x,\cdot)$ that are provided by functionals $\lambda \in \Lambda$ for which one of the constraints is active [6]. The rest of the functionals, infinite or not, is irrelevant. The "active" functionals correspond to the "support vectors" of learning machines. Numerical techniques can focus on iterations using only the currently "active" functionals, and this can be done by sequential quadratic programming. Some performance results, theoretical and numerical, will be presented at the conference, if time permits.
The introduced tolerance can reduce the complexity of the approximation of the solution dramatically. Examples will be given at the conference. But the effect seems to be hard to analyze theoretically, because it depends very much on the function u to be reconstructed. Here is a lot of leeway for future research, as the conference talk will point out.
Greedy algorithms [9, 2] provide a way to get around the sequential quadratic programming. When working on an approximation u_N based on exact interpolation of data $\lambda_1(u),\dots,\lambda_N(u)$, they select a functional $\lambda_{N+1} \in \Lambda$ for which the error $|\lambda(u-v)|$ is largest and add it to the current set, refining the calculation via a fast Cholesky update step. It turns out that such algorithms, even if compared to the true iterative solution of (3) on active sets, can be quite effective in some cases, but are slow in others. Maybe that future research on preconditioning, as treated in the next short story, will help.

3 Short Story on Scaling and Preconditioning

By a surprising empirical observation of Driscoll and Fornberg [1] smooth radial basis function interpolants tend to converge towards polynomials, when the scaling gets wide, i.e. when the systems get more and more ill–conditioned. A thorough analysis of this limit behaviour [7] should provide some information that helps to counteract the ill–conditioning. Note that even for the case of plain interpolation by Gaussians or multiquadrics there are no results so far that help to understand the limit for wide scales.
Unfortunately, this story cannot avoid some technical details, and it must currently be confined to scales of (strictly) positive definite and analytic radial basis functions of the form

$$\Phi_c(x,y) = \phi(c\|x-y\|_2) = f(c^2\|x-y\|_2^2) = \sum_{\ell=0}^{\infty} f_\ell c^{2\ell}\|x-y\|_2^{2\ell} \tag{4}$$

with functions f that are analytic around zero. Inverse multiquadrics and Gaussians are the predominant examples, but note that the scaling adopted here is such that $c \to 0$ means that the bell–shaped function $\phi(cr)$ gets wider and wider.
Let us forget about scaling for a moment and focus on given scattered points $x_1,\dots,x_M$ for interpolation. Their geometry is now related to multivariate polynomials in a specific

way. Define P_m^d as the space of d–variate polynomials of order (=degree-1) at most m and consider a basis $p_1,\ldots,p_q$ with $q=\binom{m-1+d}{d}$. Then form the $q\times M$ matrices P_m with entries $p_j(x_k)$, $1\le j\le q$, $1\le k\le M$ and look at the nested subspaces $D_m := \ker P_m \subseteq \mathbb{R}^M$ which satisfy

$$\mathbb{R}^M = D_0 \supseteq D_1 \supseteq \cdots \supseteq D_\mu \neq \{0\} = D_{\mu+1} \tag{5}$$

with a uniquely defined number μ that depends on the data points and the space dimension. With this space decomposition one can form bases of D_μ, $D_{\mu-1},\ldots$ until a full basis $\alpha^1,\ldots,\alpha^M$ of vectors in $\mathbb{R}^M$ with

$$\alpha^j \in D_{t_j}\setminus D_{t_j+1} \text{ for } 0=t_0\le t_1\le\cdots\le t_M=\mu$$

is reached. These vectors act like divided differences. In fact, they have [7] the properties

$$\begin{aligned}\sum_{k=1}^M \alpha_k^j p(x_k) &= 0 \text{ for all } p\in P_{t_j}\\ \deg\left(\sum_{k=1}^M \alpha_k^j \|x-x_k\|_2^{2\ell}\right) &\le 2\ell - t_j.\end{aligned}$$

The $M\times M$ matrix A formed by these M vectors accounts for the geometry of the points $x_1,\ldots,x_M$. The scaling of the radial basis function interpolating on these points is independently treated via diagonal matrices B_c with entries c^{-t_j}, $1\le j\le M$ on the diagonal. The scaled interpolation system has the matrix C_c with entries $\phi(c\|x_j-x_k\|_2)$, and we form the symmetric matrix $B_cAC_cA^TB_c$ which turns out to have a well–defined limit, because its (r,s)–element is

$$\begin{aligned}
& \sum_{j=1}^M c^{-t_r}\alpha_j^r \sum_{k=1}^M c^{-t_s}\alpha_k^s \phi(c\|x_j-x_k\|_2)\\
= & \sum_{\ell=0}^\infty f_\ell c^{2\ell} \sum_{j=1}^M\sum_{k=1}^M c^{-t_r}\alpha_j^r c^{-t_s}\alpha_k^s \|x_j-x_k\|_2^{2\ell}\\
= & \sum_{2\ell\ge t_r+t_s} f_\ell c^{2\ell-t_r-t_s} \sum_{j=1}^M\sum_{k=1}^M \alpha_j^r\alpha_k^s \|x_j-x_k\|_2^{2\ell}\\
\to & f_{(t_r+t_s)/2}\sum_{j=1}^M\sum_{k=1}^M \alpha_j^r\alpha_k^s \|x_j-x_k\|_2^{t_r+t_s} && \text{for } t_r+t_s \text{ even}\\
\to & 0 && \text{for } t_r+t_s \text{ odd}
\end{aligned}$$

in the limit $c\to 0$. Here we used the second property of the penultimate display twice to kill all terms with $2\ell < t_r+t_s$, and by an additional argument [7] one can prove that the resulting matrix is positive definite. The checkerboard structure is somewhat surprising, and so is the form of the preconditioning matrices B_cA. They consist of a product of the fixed matrix A depending on the geometry of the data, but not on the scaling, while the scaling is done by the diagonal matrices B_c with the strange powers c^{-t_j} that are depending on the geometry again. As a final surprise, the preconditioning matrices are independent of the radial basis function ϕ, while the resulting matrix is not.
During the conference, some examples will be shown. Current work extends this to general functionals and conditionally positive definite radial basis functions. But there are

other interesting things to be deduced from the above technique, e.g. the fact that for interpolation on $x_1, \ldots, x_M$ with μ as in (5) one can work with the terms of (4) up to degree 2μ and generate a polynomial interpolant. Preconditioning a large system by this technique is surely not intended here. The emphasis lies in understanding the basic principles of the limit process. We see that a stabilization requires a suitable scaled projection to a space of polynomials that occurs in the limit. And it will probably be this space that occurs when researchers report the fact that plain truncation of the Cholesky decomposition gives reasonable results.

References

[1] T.A. Driscoll and B. Fornberg. Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.*, to appear, 2002.

[2] Y.C. Hon, R. Schaback, and X. Zhou. Adaptive greedy algorithms for solving large RBF collocation problems. manuscript, 2001.

[3] E. J. Kansa. Application of hardy's multiquadric interpolation to hydrodynamics. In *Proc. 1986 Simul. Conf., Vol. 4*, pages 111–117, 1986.

[4] R. Schaback. Native Hilbert spaces for radial basis functions I. In M.D. Buhmann, D. H. Mache, M. Felten, and M.W. Müller, editors, *New Developments in Approximation Theory*, number 132 in International Series of Numerical Mathematics, pages 255–282. Birkhäuser Verlag, 1999.

[5] R. Schaback. A unified theory of radial basis functions (native Hilbert spaces for radial basis functions II). *J. Comp. Appl. Math.*, 121:165–177, 2000.

[6] R. Schaback. Mathematical results concerning kernel techniques. Manuscript for the 13th IFAC Symposium on System Identification – SYSID 2003, Rotterdam, The Netherlands, 2002.

[7] R. Schaback. Multivariate interpolation by polynomials and radial basis functions. Manuscript, 2002.

[8] R. Schaback. On the versatility of meshless kernel methods. Manuscript for the International Conference on Computational and Experimental Engineering and Sciences ICCES03, Corfu, Greece, 2003.

[9] R. Schaback and H. Wendland. Adaptive greedy techniques for approximate solution of large RBF systems. *Numer. Algorithms*, 24:239–254, 2000.

[10] R. Schaback and H. Wendland. Characterization and construction of radial basis functions. In N. Dyn, D. Leviatan, D.and Levin, and A. Pinkus, editors, *Multivariate Approximation and Applications*, pages 1–24. Cambridge University Press, 2001.

[11] I. J. Schoenberg. Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, 44:522–536, 1938.

[12] B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, 2002.

[13] J. Stewart. Positive definite functions and generalizations, an historical survey. *Rocky Mountain J. Math.*, 6:409–434, 1976.

[14] Z. Wu. Hermite–Birkhoff interpolation of scattered data by radial basis functions. *Approximation Theory and its Applications*, 8/2:1–10, 1992.