Memo 5: Origin of the moment conditions in RBF interpolation and some related topics

Grady Wright

October 22, 2018

Let $\Phi : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ be a radial kernel, i.e. for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $\Phi(\mathbf{x}, \mathbf{y}) = \phi(\|\mathbf{x} - \mathbf{y}\|)$ for some $\phi : \mathbb{R}^{\geq 0} \longrightarrow \mathbb{R}$, where $\| \cdot \|$ denotes the two-norm.

Definition 1. *The radial kernel ϕ is called positive definite*[1] *on $\mathbb{R}^d$ if for any distinct set of point $X = \{\mathbf{x}_j\}_{j=1}^N$, $\mathbf{x}_j \in \mathbb{R}^d$, the matrix $A_{i,j} = \phi(\|\mathbf{x}_i - \mathbf{x}_j\|)$, $i, j = 1, \ldots, N$, is positive definite, i.e. for any vector $\mathbf{c} \in \mathbb{R}^N$*

$$\mathbf{c}^T A \mathbf{c} = \sum_{i=1}^N \sum_{j=1}^N c_i \phi(\|\mathbf{x}_i - \mathbf{x}_j\|) c_j \geq 0$$

and equality of the quadratic form happens only if $\mathbf{c} \equiv \mathbf{0}$.

Positive definite radial kernels obviously lead to non-singular interpolation problems.

Definition 2. *The radial kernel ϕ is called conditionally positive definite of order $m \geq 0$*[2] *if for any distinct set of point $X = \{\mathbf{x}_j\}_{j=1}^N$ the matrix $A_{i,j} = \phi(\|\mathbf{x}_i - \mathbf{x}_j\|)$, $i, j = 1, \ldots, N$ is positive definite on the subspace $C_m \subset \mathbb{R}^N$ given as follows:*

$$C_m = \left\{ \mathbf{b} \in \mathbb{R}^N \Bigg| \sum_{j=1}^N b_j p_k(\mathbf{x}_j) = 0,\ k = 1, \ldots, L \right\}, \tag{1}$$

where $\{p_k\}_{k=0}^L$ are a basis for the space of d-variate polynomials of degree m, P_m^d, and $L = \dim(P_m^d) = \binom{m+d}{d}$. Thus, for any $\mathbf{c} \in C_m$ with $\mathbf{c} \not\equiv 0$, $\mathbf{c}^T A \mathbf{c} > 0$.

Conditionally positive radial kernels may lead to singular interpolation problems. To get around this issue, we can impose additional constraints on the interpolation problem so that the problem is well posed. The idea is to change the interpolation problem

$$A\mathbf{c} = \mathbf{f}$$

to the equality constrained quadratic programming problem:

$$\begin{gathered} \min J(\mathbf{c}) = \frac{1}{2}\mathbf{c}^T A \mathbf{c} - \mathbf{c}^T \mathbf{f} \\ \text{subject to } P^T \mathbf{c} = 0, \end{gathered} \tag{2}$$

where $P_{i,j} = p_j(\mathbf{x}_i)$, $i = 1, \ldots, N$ and $j = 1, \ldots, L$. The constraint $P^T \mathbf{c} = 0$ is equivalent to the statement that $\mathbf{c} \in C_m$ from (1). Thus, if this constraint can be satisfied then $J(\mathbf{c})$ will have a unique minimum since A is positive definite on C_m.

[1] In other texts this would be the definition for a *strictly positive definite function*, but I hate the modifier strictly and think this definition is much more natural.

[2] In other texts this would be the definition for a *strictly conditionally positive definite function of order $m - 1$*, but this is a stupid way to define these kernels.

We can solve (2) using Lagrange multipliers. To this end, consider the *Lagrangian*

$$\mathcal{L}(\mathbf{c}, \boldsymbol{\lambda}) = J(\mathbf{c}) + \boldsymbol{\lambda}^T(P^T c). \tag{3}$$

We are interested in finding the location where the gradient of $\mathcal{L}$ with respect to the elements of $\mathbf{c}$ is zero at the same time the gradient of $\mathcal{L}$ with respect to the elements of $\boldsymbol{\lambda}$ are zero (these are called the optimality constraints). This leads to the following coupled system:

$$\nabla_{\mathbf{c}}\mathcal{L}(\mathbf{c}, \boldsymbol{\lambda}) = A\mathbf{c} - \mathbf{f} + P\boldsymbol{\lambda} = 0 \quad \text{and} \quad \nabla_{\boldsymbol{\lambda}}\mathcal{L}(\mathbf{c}, \boldsymbol{\lambda}) = P^T\mathbf{c} = 0. \tag{4}$$

Putting these together gives the linear system:

$$\begin{bmatrix} A & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}. \tag{5}$$

This is the standard RBF interpolation system for a conditionally positive definite radial kernel.

A solution of (5) (equivalently (4)) $(\mathbf{c}^*, \boldsymbol{\lambda}^*)$ is called a saddle point of the Lagrangian (3) since

$$\mathcal{L}(\mathbf{c}^*, \boldsymbol{\lambda}) \leq \mathcal{L}(\mathbf{c}^*, \boldsymbol{\lambda}^*) \leq \mathcal{L}(\mathbf{c}, \boldsymbol{\lambda}^*)$$

for any $\mathbf{c} \in \mathbb{R}^N$ and $\boldsymbol{\lambda} \in \mathbb{R}^L$.

There will be a unique solution of (5) (equivalently (4)) if P has full column rank. The proof of this result can be done by *reductio ad impossibile*. Suppose that there is a non-trivial solution to the homogeneous system (5) (i.e. a solution with $\mathbf{f} = 0$). The first block equation then becomes, after multiplying by $\mathbf{c}^T$,

$$\mathbf{c}^T A\mathbf{c} + \mathbf{c}^T P\boldsymbol{\lambda} = 0.$$

Since $\mathbf{c}^T P = P^T\mathbf{c} = 0$, the previous expression simplifies to $\mathbf{c}^T A\mathbf{c} = 0$. However, A is conditionally positive definite of order m, so $\mathbf{c}^T A\mathbf{c} = 0$ necessarily implies that $\mathbf{c} \equiv 0$. The only possibility then is that $\boldsymbol{\lambda}$ is non-trivial. But, P has full column rank, so that the only solution to $P\boldsymbol{\lambda} = 0$ is $\boldsymbol{\lambda} \equiv 0$.

Note that including more polynomial terms in the constraints on the interpolation problems does not change the uniqueness of the solution, provided P is full column rank. This follows from the fact that $C_0 \supset C_1 \supset C_2 \supset \cdots$.

Since were on the topic of the saddle point system (5), let's look at an alternative way to solve it than a direct method or one based on the Schur complement. This approach is sometimes called the null-space method or reduced Hessian method or dual variable method. Suppose that the N-by-L matrix P has full column rank, and let the N-by-$(N-L)$ matrix Z contain linearly independent columns that span the left null space of P, i.e. $P^T Z = Z^T P = \mathbf{0}$. We can thus express the solution to $P^T\mathbf{c} = 0$ as $\mathbf{c} = Z\mathbf{d}$, for some new unknown $\mathbf{d}$. Plugging this into the first block equation from (5) gives

$$AZ\mathbf{d} + P\boldsymbol{\lambda} = \mathbf{f}.$$

Pre-multiplying this equation by Z^T and noting that $Z^T P = \mathbf{0}$ by construction, we have the linear system

$$Z^T AZ\mathbf{d} = Z^T\mathbf{f}.$$

The $(N-L)$-by-$(N-L)$ matrix $Z^T AZ$ is positive definite since A is conditionally positive definite of order m and the columns of Z are necessarily in the subspace C_m from (1). Once this system is solved we can recover $\mathbf{c}$ from the relation $\mathbf{c} = Z\mathbf{d}$, and then recover $\boldsymbol{\lambda}$ by computing the least squares solution to $P\boldsymbol{\lambda} = \mathbf{f} - A\mathbf{c}$, i.e. solving

$$P^T P\boldsymbol{\lambda} = P^T(\mathbf{f} - A\mathbf{c}). \tag{6}$$

An interesting question is whether the null-space approach can be adapted to the case where P does not have full column rank as in the case where the nodes $\{\mathbf{x}_j\}_{j=1}^N$ lie on some algebraic surface.

Many of the results on the linear algebra discussed here can be found in [BGL05].

References

[BGL05] Michele Benzi, Gene H Golub, and Jorg Liesen. Numerical solution of saddle point problems. *Acta numerica*, 14(1):1–137, 2005.