Supplementary Notes: Representation of Thin-Plate Splines

Advanced Topics in Statistical Learning, Spring 2024

Alden Green

Given data $(x_1, y_1), \dots, (x_n, y_n)$ with each $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$, the thin-plate spline problem is to solve

$$\underset{f}{\text{minimize}} \ \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda J_d^m(f), \tag{1}$$

where $J_d^m(f) = \sum_{|\alpha|=m} \frac{m!}{\alpha!}\int_{\mathbb{R}^d}(D^\alpha f(x))^2\,dx$ is a derivative-based penalty of roughness.

Thin-plate splines were originally proposed by Duchon (1977), though in the context of interpolation rather than penalized regression. Duchon showed that when $m > d/2$ there is a representer theorem for the thin-plate spline problem (1) that allows it to be recast as a generalized ridge regression. The result is analogous to representation for penalized regression in Reproducing Kernel Hilbert Spaces (RKHS), but with the added subtlety that the regularizer $J_d^m(f)$ has a non-trivial null space. For this reason the representer theorem for (1) does not immediately follow from results for RKHS. Instead (1) must be derived separately, and the proof must correctly handle the null space of $J_d^m(f)$.

The proof of Duchon relies heavily on Fourier transforms and is overall a little technical. Subsequent papers Meinguet (1979); Wahba and Wendelberger (1980) and the book on smoothing splines Gu (2013) give a simplified and more constructive analysis, but the first reference still requires an appreciable level of mathematical sophistication and I found that the latter two go a little fast over some of the details. (Though all three are excellent references.) In this note I explicitly state the thin-plate spline representer theorem (Theorem 1 below) and walk through the proof.

1 The theorem

The solution to (2) is made up of radial basis functions (RBFs) and polynomials. The RBFs we care about are translates of the fundamental solution (also known as a Green's function) E of the polyharmonic equation

$$(-1)^m \Delta^m E = \delta_0. \tag{2}$$

In equation (2) the operator Δ^m is the kth-iterated Laplacian – defined recursively – and δ_0 is the Dirac impulse. The fundamental solution has the explicit form

$$E(r) = c_d \begin{cases} r^{2m-d}\ln r, & \text{if } 2m > d \text{ and } d \text{ is even} \\ r^{2m-d}, & \text{otherwise.} \end{cases}$$

Here c_d is some complicated constant I won't bother writing out. The RBF at point x_i is given by $E(x_i, \cdot) := E(|x_i - \cdot|)$.

As mentioned, the null space $\mathcal{N}$ of $J_d^m(f)$ is non-trivial and in fact consists of all polynomials of degree at most $m-1$. This is am $M = \binom{m+d-1}{d}$ dimensional vector space, and we take $\phi_1, \dots, \phi_M$ to be an arbitrary basis of $\mathcal{N}$. Let $\mathbf{\Phi} \in \mathbb{R}^{n\times M}$ have entries $\mathbf{\Phi}_{ij} = \phi_j(x_i)$ and $\mathbf{E} \in \mathbb{R}^{n\times n}$ have entries $\mathbf{E}_{ij} = E(x_i, x_j)$.

Theorem 1. *Suppose $m > d/2$. Then the solution to (1) is well-defined and can be written in the form*

$$f_\lambda(x) = \sum_{\nu=1}^{M} c_\nu \phi_\nu(x) + \sum_{i=1}^{n} d_i E(x_i, x),$$

where additionally $\sum_{i=1}^{n} d_i \phi_\nu(x_i) = 0$ *for each* ν*. The optimization can be rewritten as the* generalized ridge *problem*

$$\underset{c,d}{\text{minimize}} \; \|y - \mathbf{\Phi} c - \mathbf{E} d\|^2 + \lambda d^\top \mathbf{E} d \;\; \text{subject to} \;\; \mathbf{\Phi}^\top d = 0, \tag{3}$$

which has a unique solution so long as $\operatorname{rank}(x_1, \ldots, x_n) \geq M$*.*

2 Formalizing thin-plate splines

Formally speaking, to make sense of problem (1) one needs a domain. The first thought is to use the Sobolev space $W^{m,2}(\mathbb{R}^d)$ which is the space of m-times weakly differentiable functions for which $D^\alpha f \in L^2(\mathbb{R}^d)$ for all $|\alpha| \leq m$. Technically speaking the Sobolev space is made up of equivalence classes of functions that agree up to sets of measure zero, but when $m > d/2$ each equivalence class contains a continuous representative (by the Sobolev Embedding Theorem) and so we can restrict our attention to continuous functions.

A more fundamental concern is that polynomials are not in $W^{m,2}(\mathbb{R}^d)$ and so clearly that cannot be the right space to search for the solution. This issue is cleared up by using the Beppo Levi space, which consists of functions

$$\mathrm{BL}_m(\mathbb{R}^d) = \Big\{ f \in C(\mathbb{R}^d), \quad D^\alpha f \text{ exists and } D^\alpha f \in L^2(\mathbb{R}^d) \;\text{ for all } \alpha = |m| \Big\}, \tag{4}$$

and is equipped with the semi-inner product

$$(u, v)_{\mathrm{BL}_m(\mathbb{R}^d)} := \sum_{|\alpha| = m} \frac{m!}{\alpha!} (D^\alpha u, D^\alpha v)_{L^2}. \tag{5}$$

For simplicity write $(u, v)_m = (u, v)_{\mathrm{BL}_m(\mathbb{R}^d)}$. Throughout, the minimum in (1) should be interpreted as being over all functions $f \in \mathrm{BL}_m(\mathbb{R}^d)$.

3 Road map

Let's start with a road map for the proof of Theorem 1. The fundamental idea will be to define an inner product $\langle \cdot, \cdot \rangle$ on $\mathrm{BL}_m(\mathbb{R}^d)$, allowing us to "project out" polynomials in the resulting inner product space $\mathcal{H}$. That is, we will decompose functions $f \in \mathcal{H}$ into the sum of two orthogonal parts

$$f = Pf + [I - P]f, \quad Pf \in \mathcal{H}_0, [I - P]f \in \mathcal{H}_1,$$

where P is the projection of $f \in \mathrm{BL}_m(\mathbb{R}^d)$ onto $\mathcal{N}$. This is useful because, as we will see, $(\cdot, \cdot)_m$ is an inner product in $\mathcal{H}_1 = \{[I - P]f : f \in \mathcal{H}\}$, and the resulting Hilbert space is an RKHS; letting $\{\eta_x^{(m)} : x \in \mathbb{R}^d\}$ be the representers of evaluation, we can use standard arguments in the theory of RKHS to show that $[I - P]f_\lambda \in \operatorname{span}\{\eta_{x_1}^{(m)}, \ldots, \eta_{x_n}^{(m)}\}$. This is what will end up happening in Proposition 1. Finally we will use an explicit representation of the reproducing kernel $\eta^{(m)}$ to rewrite the solution in terms of $E_{x_1}, \ldots, E_{x_n}$, giving us the desired result.

4 Direct sum RKHS

In order to execute this plan we need to define a suitable inner product on $\mathrm{BL}_m(\mathbb{R}^d)$ with which to project out polynomials. To that end let $\{z_0, \ldots, z_M\}$ be a collection of $\mathcal{N}$-unisolvent points, meaning

$$\sum_{\nu=1}^{M} c_\nu \phi_\nu(z_k) = 0 \;\text{ for all } k \Longrightarrow c = 0.$$

Now introduce a second semi-inner product over $\mathrm{BL}_m(\mathbb{R}^d)$:

$$(u,v)_0 := \sum_{j=1}^{M} u(z_j)v(z_j). \tag{6}$$

Then $\langle u,v\rangle := (u,v)_0 + (u,v)_m$ is an inner product on $\mathrm{BL}_m(\mathbb{R}^d)$, and

$$\mathcal{H} = \{f \in \mathrm{BL}_m(\mathbb{R}^d) : \langle f,f\rangle < \infty\}$$

is an RKHS.

Now we introduce the spaces $\mathcal{H}_0, \mathcal{H}_1$ alluded to above. $\mathcal{H}_0$ is simply the null space $\mathcal{N}$ of $J_m^d(\cdot)$, equipped with the inner product $(\cdot,\cdot)_0$ defined in (6), and $\mathcal{H}_1$ is the orthogonal complement:

$$\mathcal{H}_1 = \big\{f \in \mathcal{H} : f(z_k) = 0 \text{ for each } z_k\big\}.$$

These spaces are orthogonal in the sense that $\langle f_0, f_1\rangle = 0$ for $f_0 \in \mathcal{H}_0$ and $f_1 \in \mathcal{H}_1$. Additionally both are RKHS, and the reproducing kernel $\eta(x,y)$ of $\mathcal{H}$ can be decomposed as

$$\eta(x,y) = \eta^{(0)}(x,y) + \eta^{(m)}(x,y),$$

where $\eta^{(0)}$ is the reproducing kernel of $\mathcal{H}_0$ and $\eta^{(m)}$ the reproducing kernel of $\mathcal{H}_1$. So we can write $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ as a direct sum RKHS.

As mentioned, we prove Theorem 1 by establishing a representation of the solution in terms of $\eta^{(m)}$, and then rewriting this as a function of E. The relationship between $\eta^{(m)}$ and E is given in terms of the projection operator P in Theorem 2. To write things more explicitly, taking $\{\varphi_1, \ldots, \varphi_M\}$ to be an orthobasis of $\mathcal{H}_0$ we can write the projection operator as $Pf(x) = \sum_\nu (\varphi_\nu, f)_0 \varphi_\nu$, and for simplicity taking $\{\varphi_1, \ldots, \varphi_M\}$ to be the canonical basis (so that $\varphi_\nu(z_k) = \delta_{\nu k}$) this becomes

$$Pf(x) = \sum_\nu f(z_\nu)\varphi_\nu(x).$$

Theorem 2. *The reproducing kernel of $\mathcal{H}_1$, evaluated at a given $x, y \in \mathbb{R}^d$, is given by*

$$\eta^{(m)}(x,y) = E(x,y) - \sum_\nu E(x,z_\nu)\varphi_\nu(y) - \sum_{\nu'} E(z_{\nu'},y)\varphi_{\nu'}(x) + \sum_{\nu,\nu'} \varphi_\nu(x)\varphi_{\nu'}(y)E(z_\nu, z_{\nu'}). \tag{7}$$

Proof. This result and proof is due to Meinguet (1979). I will go fast without really doing the argument justice, and you should consult Meinguet (1979) for missing details.

We begin from the definition of the reproducing kernel of $\mathcal{H}_1$:

$$\big(\eta_x^{(m)}, f\big)_m = f(x), \quad \text{for all } f \in \mathcal{H}_1, x \in \mathbb{R}^d.$$

Let $\mathcal{D}$ be the Schwartz space of distributions, i.e. continuous linear functionals on the set of smooth and compactly supported functions $C_c^\infty(\mathbb{R}^d) = \mathcal{D}'$. By definition $[I-P]f \in \mathcal{H}_1$ and so

$$f(x) - \sum_{\nu=1}^{M} f(z_\nu)\varphi_\nu(x) = \big(\eta^{(m)}, [I-P]f\big)_m = \big(\eta_x^{(m)}, f\big)_m. \tag{8}$$

Now we make use of repeated application of integration by parts – keeping in mind that $f \in \mathcal{D}'$ is compactly supported – to deduce that

$$\big(\eta_x^{(m)}, f\big)_m = \big((-1)^m \Delta^m \eta_x^{(m)}, f\big),$$

where $(\cdot,\cdot)$ is the duality pairing between $\mathcal{D}$ and $\mathcal{D}'$. Rewriting the left hand side of (8) in terms of this duality pairing, we obtain that $\eta_x^{(m)} \in \mathcal{H}_1$ satisfies the differential equation

$$(-1)^m \Delta^m F_x = \delta_x - \sum_{\nu=1}^{M} \varphi_\nu(x) \delta_{z_\nu}, \tag{9}$$

which has to be interpreted in the distributional sense: both sides of (8) are operators which act in the same way on all $f \in \mathcal{D}'$.

Now from the definition of E we can easily get a distribution H_x that satisfies (9), although H_x is not in $\mathcal{H}_1$:

$$H_x = E_x - \sum_{\nu=1}^{M} \varphi_\nu(x) E_{z_\nu}.$$

At this point we will use an important result without proof: $H_x \in \mathcal{H}$. To get a sense of why this is non-trivial, note that the same statement does not hold true for E_x. However, once we take for granted that $H_x \in \mathcal{H}$ the proof is almost finished; clearly $[I-P]H_x \in \mathcal{H}_1$ and in fact $[I-P]H_x$ is the unique function in $\mathcal{H}_1$ for which (9) is satisfied (since P projects to the null space of Δ^m). So $\eta^{(m)} = [I-P]H_x$. Solving for $[I-P]H_x$ in terms of E then gives the desired result.

5 A first representer theorem for TPS

Now we have everything we need to represent f_λ in terms of the reproducing kernel of $\mathcal{H}_1$.

Proposition 1. *The solution to* (1) *can be written as*

$$f_\lambda = \sum_{\nu=1}^{M} c_\nu \phi_\nu + \sum_{i=1}^{n} d_i \eta_{x_i}^{(m)}. \tag{10}$$

Thus, letting $\mathbf{Q} \in \mathbb{R}^{n\times n}$ *have entries* $\mathbf{Q}_{ij} = \eta^{(m)}(x_i, x_j)$*, the problem* (1) *can be recast as the generalized ridge problem*

$$\underset{c,d}{\text{minimize}} \ \|y - \mathbf{\Phi} c - \mathbf{Q} d\|^2 + \lambda d^\top \mathbf{Q} d \ \text{ subject to } \ \mathbf{\Phi}^\top d = 0. \tag{11}$$

Proof. Decompose $f_\lambda = P f_\lambda + [I-P] f_\lambda$. Observe that since the smoothness functional $J_m^d(f_\lambda) = (f_\lambda, f_\lambda)_m = ([I-P]f_\lambda, [I-P]f_\lambda)$ we can rewrite the objective in (1), evaluated at its minimizer f_λ, as

$$\frac{1}{n}\sum_{i=1}^{n}\Big(y_i - Pf_\lambda(x_i) - [I-P]f_\lambda(x_i)\Big)^2 + \lambda J_d^m([I-P]f_\lambda).$$

From here the analyis is completely standard for the theory of reproducing kernels. Further decompose $[I-P]f_\lambda = f_\lambda^{(\pi)} + f_\lambda^{\perp}$, where $f_\lambda^{(\pi)} \in \text{span}(\eta_{x_i}^{(m)})$ and $f_\lambda^{\perp}$ belongs to the orthogonal complement. Since $\eta_{x_i}^{(m)}$ is the representer of evaluation for $\mathcal{H}_1$ and since $f_\lambda^{(\pi)}$ and $f_\lambda^{\perp}$ are orthogonal it follows that

$$f_\lambda^{\perp}(x_i) = (f_\lambda^{\perp}, \eta_{x_i}^{(m)})_m = 0.$$

Also, again using the orthogonality of $f_\lambda^{(\pi)}$ and $f_\lambda^{\perp}$,

$$J_d^m\big([I-P]f_\lambda\big) = J_d^m(f_\lambda^{(\pi)}) + J_d^m(f_\lambda^{\perp}) \geq J_d^m(f_\lambda^{(\pi)}),$$

with equality only if $f_\lambda^{\perp} = 0$. We conclude that $f_\lambda^{\perp} = 0$, from which (10) follows.

Finally, to complete the proof of Proposition 1 we need to show the equivalence between the original problem and (11). First of all, noting that $f_\lambda(x_i) = \mathbf{\Phi}_{i\cdot} c + \mathbf{Q}_{i\cdot} d$ and $J_d^m([I-P]f_\lambda) = d^\top \mathbf{Q} d$, we see that (1) can be rewritten as

$$\underset{c,d}{\text{minimize}} \ \|y - \mathbf{\Phi} c - \mathbf{Q} d\|^2 + \lambda d^\top \mathbf{Q} d,$$

which is (11) without the constraint $\mathbf{\Phi}^\top d = 0$. To see why this constraint additionally holds, we look at the normal equations: at the minimizer $(\hat{c}, \hat{d})$ of (11),

$$\begin{aligned}\mathbf{\Phi}^\top y &= \mathbf{\Phi}^\top \mathbf{Q}\hat{d} + \mathbf{\Phi}^\top \mathbf{\Phi}\hat{c} \\ \mathbf{Q}y &= \lambda \mathbf{Q}\hat{d} + \mathbf{Q}\mathbf{Q}\hat{d} + \mathbf{Q}\mathbf{\Phi}\hat{c}.\end{aligned} \tag{12}$$

Since $\eta^{(m)}$ is a positive definite kernel $\mathbf{Q}$ is invertible. So we may rewrite the second equation in (12) as $y = \lambda\hat{d} + \mathbf{Q}\hat{d} + \mathbf{\Phi}\hat{c}$, and plugging this into the first equation in (12) yields

$$\lambda \mathbf{\Phi}^\top \hat{d} = 0.$$

6 Representation in terms of Green's function

Lemma 2 gives the explicit relationship between $\eta^{(m)}$ and the Green's function E. Applying this to $\sum_{i=1}^n \hat{d}_i \eta_{x_i}^{(m)}$ gives the following expression:

$$\sum_{i=1}^n \hat{d}_i \eta_{x_i}^{(m)}(t) = \sum_{i=1}^n \hat{d}_i E_{x_i}(t) - \sum_{i=1}^n \sum_{\nu} \hat{d}_i E(x_i, z_\nu)\varphi_\nu(t) + \sum_{i=1}^n \hat{d}_i \sum_{\nu'} E_t(z_{\nu'})\varphi_{\nu'}(x_i) + \sum_{i=1}^n \hat{d}_i \sum_{\nu,\nu'} \varphi_\nu(t)\varphi_{\nu'}(x_i) E(z_\nu, z_{\nu'}).$$

On the right hand side the second and fourth terms are polynomial in t. On the other hand, $\hat{d}$ lies in the kernel of $\mathbf{\Phi}^\top$ and so the third term is 0. We conclude that $\sum_{i=1}^n \hat{d}_i \eta_{x_i}^{(m)} = \sum_{i=1}^n \hat{d}_i E_{x_i} + p$ for some polynomial p, and consequently for some $\widetilde{c}$,

$$f_\lambda = \sum_{\nu=1}^M \tilde{c}_\nu \phi_\nu + \sum_{i=1}^n \hat{d}_i E_{x_i}. \tag{13}$$

7 Proof of Theorem 1

Combining (13) and Proposition 1 shows that the solution to (1) is equal to

$$\underset{c,d}{\text{minimize}} \;\; \|y - \mathbf{\Phi}c - \mathbf{E}d\|^2 + \lambda\, J_d^m\left(\sum_{i=1}^n d_i E_{x_i}\right) \;\; \text{subject to} \;\; \mathbf{\Phi}^\top d = 0. \tag{14}$$

It remains only to show that the penalty in (14) has the finite-dimensional representation as written in (10). This can be seen by another application of Lemma 2: for any d such that $\mathbf{\Phi}^\top d = 0$,

$$\begin{aligned} J_d^m\left(\sum_{i=1}^n d_i E_{x_i}\right) &= J_d^m\left(\sum_{i=1}^n d_i \eta_{x_i}^{(m)}\right) \\ &= \sum_{i,j=1}^n d_i d_j \eta^{(m)}(x_i, x_j) \\ &= \sum_{i,j=1}^n d_i d_j \left\{ E(x_i, x_j) - \sum_{\nu} E(x_i, z_\nu)\varphi_\nu(x_j) - \sum_{\nu'} E(z_{\nu'}, x_j)\varphi_{\nu'}(x_i) + \sum_{\nu,\nu'} \varphi_\nu(x_i)\varphi_{\nu'}(x_j) E(z_\nu, z_{\nu'}) \right\} \\ &= \sum_{i,j=1}^n d_i d_j E(x_i, x_j), \end{aligned}$$

where the first equality follows since $\eta_{x_i}^m$ and E_{x_i} differ by a polynomial.

References

Jean Duchon. Splines minimizing rotation-invariant semi-norms in sobolev spaces. In *Constructive Theory of Functions of Several Variables: Proceedings of a Conference Held at Oberwolfach April 25–May 1, 1976*, pages 85–100. Springer, 1977.

Chong Gu. *Smoothing spline ANOVA models*, volume 297. Springer, 2013.

Jean Meinguet. Multivariate interpolation at arbitrary points made simple. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, 30(2):292–304, 1979.

Grace Wahba and James Wendelberger. Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly weather review*, 108(8):1122–1143, 1980.