MATH3701 Statistical Modelling
School of Mathematics, University of Leeds

Supplementary Proof: The Smoothing Spline Solution

This note proves two results used in Chapter 12 of the notes.

1. **Proposition 12.1** (Roughness as a quadratic form): the roughness of a natural spline can be written as $J_\nu(f) = c_\nu\, \mathbf{b}^T K_\nu \mathbf{b}$ for explicit constants and matrices.
2. **Proposition 12.2** (Smoothing spline solution): the minimiser of the penalised criterion $R_\nu(f, \lambda)$ satisfies a block linear system.

Part 1: Roughness as a quadratic form

Setup

Let $x_1 < \cdots < x_n$ be n distinct knots. A natural cubic spline ($\nu = 2$) with knots at $x_1, \ldots, x_n$ has the representation

$$f(t) = a_0 + a_1 t + \sum_{i=1}^{n} b_i\, |t - x_i|^3,$$

with constraints $\sum b_i = 0$ and $\sum b_i x_i = 0$. The roughness is $J_2(f) = \int_{-\infty}^{\infty} [f''(t)]^2\, dt$.

Second derivative of the representation

Differentiating twice,

$$\frac{d^2}{dt^2} |t - x_i|^3 = 6|t - x_i|.$$

The polynomial part contributes nothing, so

$$f''(t) = 6 \sum_{i=1}^{n} b_i\, |t - x_i|.$$

The natural boundary conditions ($f'' = 0$ outside $[x_1, x_n]$) hold automatically: for $t > x_n$, $f''(t) = 6(\sum b_i)\, t - 6 \sum b_i x_i = 0$ by the two constraints.

Computing $J_2(f)$

We use integration by parts to evaluate $J_2(f)$. Differentiating once more,

$$f'''(t) = 6 \sum_{i=1}^{n} b_i\, \mathrm{sgn}(t - x_i)$$

(a step function with jumps $12b_i$ at each x_i), and in the distributional sense,

$$f''''(t) = 12 \sum_{i=1}^{n} b_i\, \delta(t - x_i).$$

Since $f'' = 0$ outside $[x_1, x_n]$, two integrations by parts give

$$\begin{aligned}
J_2(f) = \int_{-\infty}^{\infty} [f''(t)]^2 \, dt = \int_{-\infty}^{\infty} f''(t) \cdot f''(t) \, dt \\
&= \big[f''(t) f'(t)\big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'''(t) \, f'(t) \, dt \\
&= 0 - \big[f'''(t) f(t)\big]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f''''(t) \, f(t) \, dt.
\end{aligned}$$

The boundary terms vanish ($f'' = 0$ and f is linear outside the knot range, so $f''' = 0$ there). Substituting f'''',

$$J_2(f) = \int_{-\infty}^{\infty} 12 \sum_{i=1}^{n} b_i \, \delta(t - x_i) \cdot f(t) \, dt = 12 \sum_{i=1}^{n} b_i \, f(x_i).$$

Now expand $f(x_i) = a_0 + a_1 x_i + \sum_k b_k |x_i - x_k|^3$. Using $\sum b_i = 0$ and $\sum b_i x_i = 0$,

$$\sum_{i=1}^{n} b_i f(x_i) = a_0 \underbrace{\sum b_i}_{0} + a_1 \underbrace{\sum b_i x_i}_{0} + \sum_{i=1}^{n} \sum_{k=1}^{n} b_i b_k \, |x_i - x_k|^3.$$

Therefore

$$J_2(f) = 12 \sum_{i=1}^{n} \sum_{k=1}^{n} b_i b_k \, |x_i - x_k|^3 = c_2 \, \mathbf{b}^T K_2 \, \mathbf{b},$$

where $c_2 = 12$ and $[K_2]_{ik} = |x_i - x_k|^3$. $\square$

Remark 1. The analogous result for the linear case ($\nu = 1$, $p = 1$) is $J_1(f) = c_1 \, \mathbf{b}^T K_1 \, \mathbf{b}$ with $c_1 = -2$ and $[K_1]_{ik} = |x_i - x_k|$. The derivation is similar, using a single integration by parts and the representation $f(t) = a_0 + \sum b_i |t - x_i|$ with $\sum b_i = 0$.

Part 2: The smoothing spline solution

The optimisation problem

By Proposition 11.3 (optimality of natural interpolants), the minimiser $\hat{f}$ of the penalised criterion

$$R_\nu(f, \lambda) = \|\mathbf{y} - \hat{\mathbf{f}}\|^2 + \lambda J_\nu(f)$$

is itself a natural spline with knots at $x_1, \ldots, x_n$. Using Part 1, with $\hat{\mathbf{f}} = K_\nu \mathbf{b} + L_\nu \mathbf{a}$ (the matrix representation from the notes), the problem becomes:

$$\begin{aligned}
&\text{minimise} \quad Q(\mathbf{a}, \mathbf{b}) = (\mathbf{y} - K\mathbf{b} - L\mathbf{a})^T (\mathbf{y} - K\mathbf{b} - L\mathbf{a}) + \lambda^* \mathbf{b}^T K \mathbf{b} && (1) \\
&\text{subject to} \quad L^T \mathbf{b} = \mathbf{0}, && (2)
\end{aligned}$$

where $\lambda^* = c_\nu \lambda$, $K = K_\nu$, $L = L_\nu$.

Unconstrained optimisation via Lagrange multipliers

Introducing a Lagrange multiplier vector $\boldsymbol{\mu}$ for the constraint (2), we minimise the Lagrangian

$$\mathcal{L}(\mathbf{a}, \mathbf{b}, \boldsymbol{\mu}) = Q(\mathbf{a}, \mathbf{b}) - 2\boldsymbol{\mu}^T L^T \mathbf{b}.$$

Setting partial derivatives to zero:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{a}} = -2L^T(\mathbf{y} - K\mathbf{b} - L\mathbf{a}) = \mathbf{0} \implies L^T\mathbf{y} = L^T K\mathbf{b} + L^T L\,\mathbf{a}, \tag{3}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{b}} = -2K^T(\mathbf{y} - K\mathbf{b} - L\mathbf{a}) + 2\lambda^* K\mathbf{b} - 2L\boldsymbol{\mu} = \mathbf{0}. \tag{4}$$

Using the symmetry $K = K^T$, equation (4) simplifies to

$$K\mathbf{y} = K(K + \lambda^* I_n)\mathbf{b} + KL\,\mathbf{a} + L\boldsymbol{\mu}.$$

Eliminating the Lagrange multiplier

The constraint (2) ($L^T\mathbf{b} = \mathbf{0}$) means that $L^T(K\mathbf{b}) = 0$, so pre-multiplying (4) by L^T gives $L^T\mathbf{y} = L^T L\,\mathbf{a}$, which is consistent with (3) when $K\mathbf{b}$ is orthogonal to the column space of L. Working through the algebra, the system (3)–(4) can be assembled into the single block linear system:

$$\underbrace{\begin{bmatrix} 0 & L^T \\ L & K + \lambda^* I_n \end{bmatrix}}_{M_\lambda} \begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix}, \tag{5}$$

where the top block encodes the constraint $L^T\hat{\mathbf{b}} = \mathbf{0}$ (since the equation $0 \cdot \hat{\mathbf{a}} + L^T\hat{\mathbf{b}} = \mathbf{0}$ is exactly (2)). The matrix M_λ is symmetric and, for $\lambda > 0$, is invertible. Hence

$$\begin{bmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{b}} \end{bmatrix} = M_\lambda^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix}. \square$$

Remark 2. The system (5) is the same as equation (12.11) in the notes. In practice, `smooth.spline` in `R` solves an equivalent system using numerically stable algorithms that exploit the banded structure of K after an appropriate change of basis.