Problem 0. *The Smoothing Spline Problem*
Let f be a natural cubic spline defined on the interval $[a,\, b]$, interpolating values $y_1, \ldots, y_n$ at the n knots $x_1 \ldots, x_n$, where $a < x_1 < \ldots < x_n < b$. Let $\tilde{f}$ be any continuous twice differentiable function on $[a,\, b]$ such that $\tilde{f}(x_i) = y_i$ for $i = 1, \ldots, n$. Put $h = \tilde{f} - f$.

(i) Show that

$$\int_a^b f''(x)h''(x)dx = 0\,.$$

(ii) Deduce that

$$\int_a^b |\tilde{f}''(x)|^2 dx \geq \int_a^b |f''(x)|^2 dx\,.$$

(iii) When does equality hold in *(ii)*? Conclude that the natural cubic spline has the minimum value of $\int |f''(x)|^2 dx$ amongst all smooth curves that interpolate the data $\mathcal{L}_n = \{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}$.

(iv) Recall the smoothing spline optimisation problem.

(v) Argue that the solution to the smoothing spline problem in *(iv)* is necessarily a natural cubic spline.

Problem 1. *Basis Function for Natural Cubic Splines*
Consider a cubic splines f with K interior knots $x_1, \ldots, x_K$,

$$f(x) = \sum_{j=0}^{3} \beta_j\, x^j + \sum_{k=1}^{K} \lambda_k\, (x - x_k)_+^3\,,$$

where $(x)_+ := \max(0, x)$.

(i) Prove that the natural boundary conditions for natural cubic splines imply the following linear constraints on the coefficient:

$$\beta_2 = 0\,, \qquad \beta_3 = 0\,, \qquad \sum_{k=1}^{K} \lambda_k = 0\,, \qquad \sum_{k=1}^{K} x_k \lambda_k = 0\,.$$

(ii) Using *(i)*, derive the basis function

$$g_1(x) = 1\,, \qquad g_2(x) = x\,, \qquad g_{k+2}(x) = d_k(x) - d_{K-1}(x)\,, \tag{1}$$

where

$$d_k(x) = \frac{(x - x_k)_+^3 - (x - x_K)_+^3}{x_K - x_k}\,,$$

for $k = 1, \ldots, K-2$.

Problem 2.
Making use of the basis function (1), the solution to the smoothing spline problem for a training set of size n can be written

$$f(x) = \sum_{j=1}^{n} \beta_j \, f_j(x) \,.$$

Let $\beta := (\beta_1, \ldots, \beta_n)^t$.

(i) Given a training sample $\mathcal{L}_n = \{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}$, show that the penalised criteria

$$RSS(f, \lambda) = \sum_{j=1}^{n} (y_j - f(x_j))^2 + \lambda \int |f''(x)|^2 dx$$

can be rewritten in matrix form

$$RSS(\beta, \lambda) = ||y - \mathbf{W}\beta||^2 + \lambda \beta^t \, \mathbf{\Lambda} \, \beta \,,$$

for $y := (y_1, \ldots, y_n)^t$, and for some matrices $\mathbf{W}$ and $\mathbf{\Lambda}$ that you will make explicit.

(ii) Solve the optimisation problem

$$\hat{\beta} = \arg\min_{\beta} RSS(\beta, \lambda) \,,$$

and show that the estimate $\hat{y} := \mathbf{W}\hat{\beta}$ can be written $\hat{y} = \mathcal{S}_\lambda y$, where $\mathcal{S}_\lambda$ can be put into the Reinsch form $(I + \lambda \mathbf{K})^{-1}$.

Problem 3. *Smoothing Splines with tie values*

(i) We fit a model $f_\beta(x)$ parametrised by β to the learning sample $\mathcal{L}_n = \{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}$, using least squares. Show that if there are observations with tied or identical values of x, then the fit can be obtained from a reduced weighted least squares problem.

(ii) Characterize the solution to the following problem

$$\min_f \left\{ \sum_{i=1}^{n} \omega_i (y_i - f(x_i))^2 + \lambda \int_a^b |g''(x)|^2 dx \right\} ,$$

where the $\omega_i \geq 0$ are observation weights.

(iii) Deduce from *(i)* and *(ii)* the solution to the smoothing spline problem when the data have ties in x.

Problem 4. *Strictly Diagonally Dominant Matrices*
Show that a square strictly diagonally dominant matrix is invertible.