

1 Fitting the Reflux Data Numerically Using a Collocation Method¹

The Reflux ODE model given in Equation ?? can be written in the form:

$$\frac{dy}{dt} = f(y(t), t; \beta)$$

In this section, this model will be fitted parametrically, but by approximating $y(t)$ by basis functions.

Divide the interval of interest $[0, T]$ into knots $t_0 = 0 < t_1 < \dots < t_N = T$, and require that the observation points are included amongst the knots.

Assume that over each knot interval $[t_i, t_{i+1}]$ the function $y(t)$ can be approximated by a quadratic function $q_i(t)$. Furthermore, assume for the time being that the values of $y(t)$ are known at all the knot points, not just the knot points for which empirical observations are available. Let $y_i = y(t_i)$ and impose the following collocation conditions on each $q_i(t)$:

$$\begin{aligned} q(t_i) &= y_i \\ q'(t_i) &= f(y_i, t_i; \beta) \\ q'(t_{i+1}) &= f(y_{i+1}, t_{i+1}; \beta) \end{aligned}$$

Write $q_i(t)$ in the form:

$$q_i(t) = a_i(t - t_i)^2 + b_i(t - t_i) + c_i$$

Letting $h_i = t_{i+1} - t_i$, the collocation conditions then become:

$$\begin{aligned} c_i &= y_i \\ b_i &= f(y_i, t_i; \beta) \\ 2a_i h_i + b_i &= f(y_{i+1}, t_{i+1}; \beta) \end{aligned}$$

These equations can be solved for a_i, b_i , and c_i :

$$\begin{aligned} a_i &= \frac{1}{2h_i} [f(y_{i+1}, t_{i+1}; \beta) - f(y_i, t_i; \beta)] \\ b_i &= f(y_i, t_i; \beta) \\ c_i &= y_i \end{aligned}$$

Evaluating $q_i(t)$ at t_{i+1} yields that:

¹This section borrows very heavily from https://en.wikipedia.org/wiki/Collocation_method

$$\begin{aligned}
y_{i+1} &\approx q_i(t_{i+1}) \\
&= \frac{1}{2h_i} [f(y_{i+1}, t_{i+1}; \beta) - f(y_i, t_i; \beta)] h_i^2 + f(y_i, t_i; \beta) h_i + y_i \\
&= y_i + \frac{h_i}{2} [f(y_{i+1}, t_{i+1}; \beta) + f(y_i, t_i; \beta)]
\end{aligned}$$

Let S denote the set of indices for which there is an empirical observation. The discussion so far suggests the Reflux data can be fitted solving the following optimisation problem:

$$\begin{aligned}
\text{minimise:} \quad & H(\beta) = \sum_{i \in S} [y_i - \hat{y}_i]^2 \\
\text{subject to:} \quad & \hat{y}_{i+1} = \hat{y}_i + \frac{h_i}{2} [f(\hat{y}_{i+1}, t_{i+1}; \beta) + f(\hat{y}_i, t_i; \beta)]
\end{aligned}$$

While useful for illustrating more sophisticated FDA methods, the fitting methodology described here was found to not perform very well in practice. Figure 1 for example plots the result of smoothing the Melanoma data with a second derivative penalty approximated using a finite difference method. The results are poor.



Figure 1: Smoothing the Melanoma data using a finite difference approximation does not produce a particularly smooth fit!