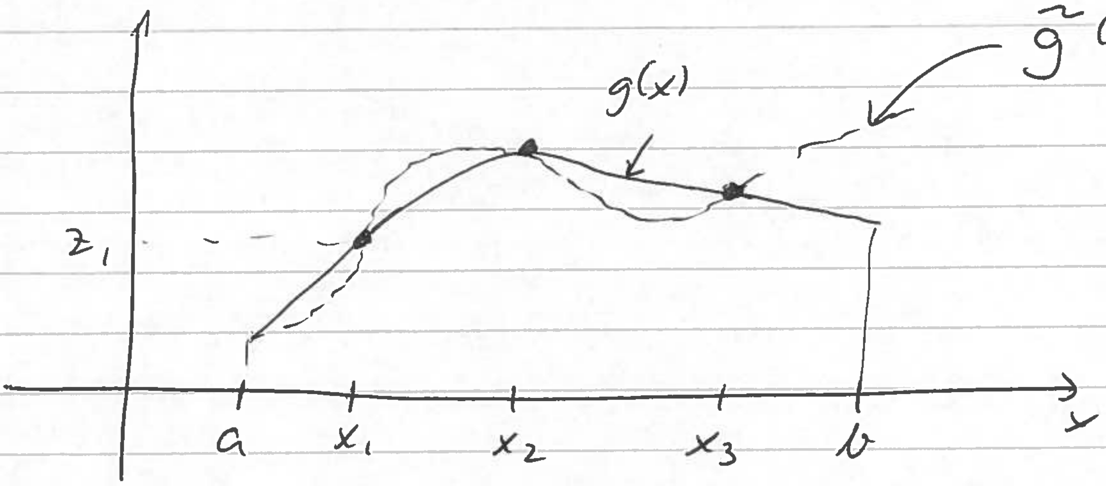
SOLUTION TO EX. 5.7

Let $g(x)$ be the natural cubic spline that goes through the points

(x_i, z_i) ; $i = 1, \dots, N$ where

$a < x_1 < \dots < x_N < b$.

This is a natural cubic spline with a knot at every x_i, which is linear in the intervals $[a, x_1]$ and $[x_N, b]$. We know that this has dimension N, so it is uniquely given because it has to go through N given points.



(a proof of the above can be found in the book of Green and Silverman – we take the existence ~~of~~ and uniqueness of such a $g(x)$ for a given fact

Then let $\tilde{g}(x)$ be any other differentiable function on $[a, b]$ that goes through the same points

(a) $h(x) = g^2(x) - g(x)$

Integration by parts: Recall

$$\int_a^b u\,dv = \Big|_a^b uv - \int_a^b v\,du$$

Here:

$dv = h''(x)dx \Rightarrow v = h'(x)$

$u = g''(x) \Rightarrow du = g'''(x)dx$

$$\int_a^b g''(x)\,h''(x)\,dx$$

$$= \underbrace{\Big|_a^b g''(x)h'(x)}_{\underbrace{g''(b)}_{=0}h'(b) - \underbrace{g''(a)}_{=0}h'(a)} - \int_a^b h'(x)g'''(x)\,dx$$

$= 0$

because g is linear outside ~~a and b~~ at a and b.

$$= -\int_a^b h'(x)\,g'''(x)\,dx$$

Here $g'''(x)$ is 0 ~~outside~~ at x_1 and x_N, and is constant in each interval.

$$= -\sum_{j=2}^{N-1} \int_{x_j}^{x_{j+1}} h'(x)\,g'''(x_j^+)\,dx$$

$$= -\sum_{j=2}^{N-1} g'''(x_j^+) \int_{x_j}^{x_{j+1}} h'(x)\,dx$$

$$= -\sum_{j=2}^{N-1} g'''(x_j^+) \left(\underbrace{h(x_{j+1}) - h(x_j)}_{\text{each of this is } 0} \right) = \underline{\underline{0}}$$

because $\tilde{g}(x_j) = g(x_j)$ for each j

(b) To prove this (see Exercise), we start by

$$0 \overset{(\ddagger)}{\leq} \int_a^b \left(\tilde{g}''(x) - g''(x)\right)^2 dx$$

because the integrand is ≥ 0

$$= \int_a^b \left(\tilde{g}''(x)\right)^2 dx - 2\int_a^b \tilde{g}''(x) g''(x)\,dx + \int_a^b \left(g''(x)\right)^2 dx \qquad (*)$$

But (a) implies that

$$\int_a^b g''(x) \left(\tilde{g}''(x) - g''(x)\right) dx = 0$$

so $$\int_a^b g''(x)\tilde{g}''(x)\,dx = \int_a^b \left(g''(x)\right)^2 dx$$

Putting this into (*) we get

$$0 \leq \int_a^b (\tilde{g}''(x))^2 dx - \int_a^b (g''(x))^2 dx$$

so $$\int_a^b (\tilde{g}''(x))^2 dx \geq \int_a^b (g''(x))^2 dx \quad (**)$$

which is what we were asked to prove.

[Note that this says that the natural cubic spline $g(x)$ minimizes $\int_a (\tilde{g}''(x))^2 dx$ among all differentiable functions on $[a,b]$ going through the points (x_i, z_i).]

We also need to show uniqueness, so that equality in (**) holds if and only if

$$\tilde{g}(x) = g(x) \text{ for \underline{all} } x.$$

From $(\ddagger)^{(\text{page } 3)}$ we see that equality in (**) holds if and only if $\tilde{g}''(x) = g''(x)$ for all x

But $\tilde{g}''(x) = g''(x)$ for all x

$$\Updownarrow$$

$\tilde{g}'(x) = g'(x) + c$ for all x

$$\Updownarrow$$

$\tilde{g}(x) = g(x) + cx + d$ for all x

But $\tilde{g}(x) = g(x)$ at N points, so we must have $c=0$, $d=0$.

and uniqueness follows.

(c) Suppose that the minimizer is $\tilde{f}$, which is not necessarily a natural cubic spline. Then let $\hat{f}$ be the natural cubic spline with knots at $x_1, \dots, x_N$ and is such that

$$\hat{f}(x_i) = \tilde{f}(x_i) \text{ for } i = 1, \dots, N$$

(the existence and uniqueness of $\hat{f}$ is such as discussed in the beginning of the solution).

Then $\hat{f}$ has the same $\sum_{i=1}^{N} (y_i - f(x_i))^2$ as $\tilde{f}$, but by (b) is

$$\int_a^b \hat{f}''(t)^2 \, dt \le \int_a^b \tilde{f}''(t)^2 \, dt \quad (\ast\ast\ast)$$

But then, since $\tilde{f}$ was assumed to be the minimizer, we must have equality in (***), and by the uniqueness shown in (b), $\tilde{f}$ must be a natural cubic spline.