

# 1 Modifying the Data2LD.opt Routine

In Section ??, for fitting the Melanoma data, the following penalised regression model is used:<sup>1</sup>

$$PENSSSE(f|\theta, \rho) = (1 - \rho) \sum [y_i - f(t_i)]^2 + (1 - \rho)\lambda \int |f - \theta f^{(4)}|^2 dt$$

While **Data2LD** can fit a penalty of the form  $f - \theta f^{(4)}$ , and it can enforce linear equality constraints such as  $\theta_1 - \theta_2 = 0$ , it cannot enforce linear inequality constraints such as  $\theta \geq 0$ , nor can it work with penalties of the form  $f - \theta^2 f^{(4)}$ .<sup>[?]</sup> This means that it's possible in principle that the estimated value of  $\theta$  returned by **Data2LD.opt** is negative even though this value is not valid for the modelling problem at hand since it is desired that  $\theta \geq 0$ .

**Approximating the Derivative:** We modified a single line in the code for the **Data2LD.opt** routine introduced in Section ?? to use a finite difference approximation to the first derivative of the objective function.

$$f'(\theta) \approx \frac{f(\theta_n) - f(\theta_{n-1})}{\theta_n - \theta_{n-1}} \quad (1)$$

Because only a single line was changed, **Data2LD.opt** still computes the second derivatives and makes use of them in the same fashion as before. Note as well that this modified code will only work for problems with only one parameter to be estimated because it will only compute the derivative in one direction.

Ignoring the line search, the optimisation routine thus takes the following form now:

$$\begin{cases} d_n &= \frac{f(\theta_n) - f(\theta_{n-1})}{\theta_n - \theta_{n-1}} \\ \theta_{n+1} &= \theta_n - d_n / f''(\theta_n) \end{cases} \quad (2)$$

The first derivative is approximated using a secant equation, but the second derivative is computed exactly.

**Error Analysis:** An error analysis must be done from scratch, as a result of the unusual situation where there are errors in the gradient, but the second derivative can be found exactly.

Suppose that instead of the true derivative  $f'(\theta)$ , the value  $f'(\theta) + \epsilon(\theta)$  is used instead. Assume that there exists  $\bar{\epsilon}_n > 0$  such that for the  $n$ th iteration,  $|\epsilon(\theta)| < \bar{\epsilon}$  for all  $\theta$  in some neighbourhood of the optimal point  $\theta^*$ . Let  $e_n = |\theta_n - \theta^*|$  denote the absolute error on the  $n$ th iteration. Then it is the case that:<sup>2</sup>

---

<sup>1</sup>In this section, for the sake of consistency with the notation used by **Data2LD**,  $\theta$  is used to denote the frequency parameter instead of the usual  $\omega$ , and  $\rho$  is used to determine the trade-off between fit to the data and fit to the ODE model instead of the  $\lambda$ .

<sup>2</sup>This result is a special case of Theorem 5.4.1 from [?].

$$e_{n+1} \leq K(\bar{e}_n + e_n^2) \quad (3)$$

To determine  $\bar{e}_n$ , it is necessary to estimate the error introduced by the approximation (1). It is straightforward<sup>3</sup> to show that the absolute value of the error is approximately equal to  $\frac{1}{2}f''(\eta)|\theta_n - \theta_{n-1}|$ . If it is further assumed that the second derivative is continuous and positive around  $\theta^*$ ,<sup>4</sup> and the sequence is sufficiently close to  $\theta^*$ , then there are constants  $c$  and  $C$  such that  $0 < cf''(\theta^*) \leq f''(\eta) \leq Cf''(\theta^*)$ .

Combining this with the Triangle Inequality, it is possible to find an upper bound that can be expressed in terms of  $e_n$  and  $e_{n-1}$ :

$$\begin{aligned} f''(\theta^*)|\theta_n - \theta_{n-1}| &\leq Cf''(\theta^*)[|\theta_n - \theta^*| + |\theta_{n-1} - \theta^*|] \\ &= Cf''(\theta^*)[e_n + e_{n-1}] \end{aligned}$$

Plugging this particular choice of  $\bar{e}_n$  into (3) yields:

$$e_{n+1} \leq K\{Cf''(\theta^*)[e_n + e_{n-1}] + e_n^2\}$$

Neglecting the quadratic  $e_n^2$  term, and letting  $M = KCf''(\theta^*)$ , we get that:

$$e_{n+1} \leq M(e_n + e_{n-1})$$

Since  $e_n \geq 0$  for all  $n$  and  $M > 0$ , this recurrence inequality implies that  $e_n \leq a_n$  for all  $n$ , where  $a_{n+1} = M(a_n + a_{n-1})$ ,  $a_0 = e_0$  and  $a_1 = e_1$ . The method thus converges linearly (in the extended sense) provided  $M$  is not too big and provided the starting position is not too far away from the optimal point.

**Comparison with the Secant Method and Successive Parabolic Interpolation:** This method only converges linearly while Successive Parabolic interpolation and the usual secant method both converge superlinearly. A crude explanation for this behaviour is that the Secant method and Parabolic Interpolation methods both effectively use three points to construct a quadratic approximation, while this method only uses two points to construct a linear approximation. While an exact second derivative is used, the gains from this are nonetheless dwarfed by the error introduced by approximating the first derivative with two points.

For the Secant Method, one finds that  $e_{n+1} \leq \bar{M}e_n e_{n-1}$ . For successive parabolic interpolation, one ends up with a recurrence relation of the form  $e_{n+1} \leq \tilde{M}[e_n e_{n-1} + e_n e_{n-2} + e_{n-1} e_{n-2}]$ . Both recurrence inequalities imply superlinear convergence subject to some mild technical conditions. [?]

<sup>3</sup>Using Taylor's theorem, it can be seen that  $f(\theta_n) - f(\theta_{n-1}) = f(\theta_n) - [f(\theta_n) - f'(\theta_n)(\theta_n - \theta_{n-1}) + \frac{1}{2}f''(\eta)(\theta_n - \theta_{n-1})^2] = f'(\theta_n)(\theta_n - \theta_{n-1}) - \frac{1}{2}f''(\eta)(\theta_n - \theta_{n-1})^2$ . Divide across by  $(\theta_n - \theta_{n-1})$  and it is then easy to get the desired result. The  $-f'(\theta_n)(\theta_n - \theta_{n-1})$  term arises because  $f(\theta_{n-1}) = f(\theta_n - (\theta_n - \theta_{n-1})) \approx f(\theta_n) - f'(\theta_n)(\theta_n - \theta_{n-1})$ .

<sup>4</sup>If this assumption fails, there are bigger problems than the rate of convergence since  $f''(\theta^*) < 0$  implies that  $\theta^*$  is not a local minimiser of the objective function.

**Comparing the Modified Method with the Original Method:** The fitting algorithm used by `Data2LD.opt` is hierarchical because there are two levels of optimisation, an outer level that computes search directions, and an inner level that conducts line searches. This makes charting the course of the method a little tricky. The easiest approach is to simply reproduce verbatim the output of `Data2LD.opt`.

Below the output for the original method and the modified method when applied to fitting the Melanoma data are presented. The workings of the two different levels can be seen. At the top level, `Data2LD.opt` computes the value of the objective function and the magnitude of the gradient used for the search direction. It then reports the values of the objective function and directional derivatives along the search direction computed in the course of conducting a line search. For the modified method, the computed gradients and directions are approximate of course.

Both methods achieve a similar value of the objective function, but the estimated values of  $\theta$  differ somewhat. The modified method is clearly much slower than the original, which is congruent with our theoretical analysis. The associated values of the objective function are so similar it is difficult to conclusively say which computed estimate is better, the fact that both are negative suggests neither should be taken too seriously.

### Newton Method with Gradient Line Search

Iter.	Criterion	Grad Length
0	0.03974	0.002238
	theta = 0.400000, dtheta = -0.001867	
	theta = -0.104395, dtheta = 0.000476	
	theta = 0.002903, dtheta = -0.000044	
1	0.039187	4.4e-05
	theta = -0.494194, dtheta = 0.002106	
	theta = -0.006478, dtheta = 0.000002	
2	0.039187	2e-06

Convergence reached.

### Newton Method with Secant Approximation To First Derivative

Iter.	Criterion	Grad Length
0	0.03974	0.002238
	theta = 0.400000, dtheta = 0.001898	
	theta = 0.414612, dtheta = -0.000300	
	theta = 0.401771, dtheta = 0.000230	
	theta = 0.402048, dtheta = -0.000005	
1	0.039554	0.001876
	theta = 0.304096, dtheta = 0.001555	
	theta = 0.318509, dtheta = -0.000241	
	theta = 0.305876, dtheta = 0.000184	
2	0.039404	0.001472
	theta = 0.209704, dtheta = 0.001177	
	theta = 0.224023, dtheta = -0.000176	
	theta = 0.211530, dtheta = 0.000134	
3	0.039293	0.001047
	theta = 0.117183, dtheta = 0.000772	
	theta = 0.131623, dtheta = -0.000107	
	theta = 0.119149, dtheta = 0.000081	
4	0.039221	0.000609
	theta = 0.026769, dtheta = 0.000353	
	theta = 0.042419, dtheta = -0.000035	
5	0.039191	0.000237
	theta = -0.034310, dtheta = 0.000030	
	theta = 0.005845, dtheta = 0.000061	
	theta = 0.010409, dtheta = -0.000007	
6	0.039187	8.1e-05
	theta = -0.021601, dtheta = -0.000015	
	theta = -0.005596, dtheta = 0.000053	
	theta = -0.000575, dtheta = 0.000002	
7	0.039187	2.7e-05

Convergence reached.