The mathematics behind straightening and flattening COVID-19 curves, and the concept of exponential period

Peter H. van der Kamp

June 27, 2020

1 Introduction

As part of the assessment for the subject MAT5OPT at La Trobe university, students have investigated COVID-19 data using the least squares method. One student introduced the notion "exponential period", i.e. the number of days that the growth is exponential ($y = c_1 \exp(c_2 x)$) before it slows down [1]. The following periods were obtained in [1], see Figure 1. A few questions arise, as one observes

	C1	C2	Begin Exponential	End Exponential	Days Exponential	Root Mean Square Error
Australia	0.0119	0.191	26 January	26 March	61	331.3939
Brazil	80.131	0.069	26 February	14 May	79	12649.96
Italy	6.379	0.151	31 January	20 March	51	6371.744
Russia	64.902	0.073	31 January	14 May	105	34136.61
Singapore	0.184	0.118	23 January	24 April	94	2074.602
South Korea	0.164	0.250	22 Janurary	2 March	41	643.1218
Sweeden	8.369	0.089	31 January	10 April	71	1802.255
United Kingdom	4.539	0.098	31 January	26 March	66	1137.219

Figure 1: Exponential periods for several countries found in [1]

that the exponential period of Italy is shorter than the one for Australia, and, that the exponent c_2 is shorter for countries that are worse off, such as Russia. In [1], exponential periods were obtained by comparing a scatter plot of the data with a least squares fitted exponential curve, and observing (with the eye) whether it fits well. For example, in Figure 2 the curve for Brazil fits well, but the curves for Sweden and Australia do not fit well (Images taken from [1]).

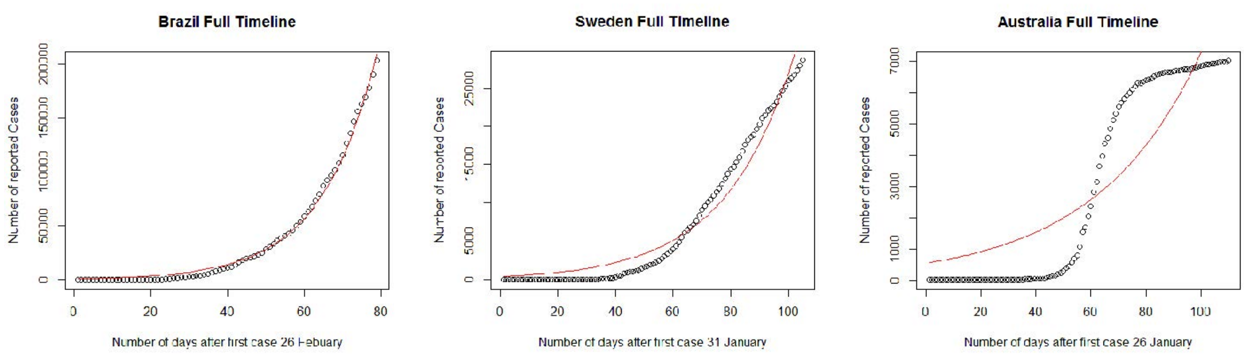

Figure 2: Exponential curves do fit or don't fit the data well, cf. [1]

The concept of exponential period is useful for the following reasons:

- it is a single number
- it is easy to understand
- it can be used to compare countries
- it is a measure of how effective lock-down measures have been (and how quick they were introduced)

In the next section we provide a method to obtain the exponential period which does not involve the eye. We consider COVID-19 data for the number of confirmed cases in Russia. This method shows that after roughly 54 days the curve is better fitted by a linear model.

In the final section, we discuss refinements and introduce two more periods or *growth phases*, namely the linear phase, and the logarithmic phase. On expects the pandemic to start off in exponential phase. One then aims to transition into a linear phase, and, with sufficient lock-down measures in place, into logarithmic phase, before eliminating the disease. Of course, it is possible to have set-backs, a second (third etc.) wave: several switches between the different types of growth are likely to occur.

2 The exponential period for COVID-19 in Russia

We break up the date for Russia, obtained from [2], into 13 periods of 9 days and fit the data for each period with both a linear model

$$y = f_{\mathbf{a}}(x) = a_1 + a_2 x,$$

and with an exponential model

$$y = f_{\mathbf{b}}(x) = b_1 \exp(b_2 x).$$

The parameters are obtained by optimising the corresponding least squares objective functions

$$F_k^{lin}(\mathbf{a}) = \sum_{i=1}^{9} \left(y_{10k+i} - f_a(x_{10k+i})\right)^2, \qquad F_k^{exp}(\mathbf{b}) = \sum_{i=1}^{9} \left(y_{10k+i} - f_b(x_{10k+i})\right)^2,$$

using linear algebra in the first case, and a downhill simplex method in the second case. The parameters we have obtained, for the data of confirmed cases in Russia from 28/2/2020 till 23/06/2020, are given in the following table: Graphs of the data and fitted curves for the first 9 periods are given in Figure

period	a_1	a_2	b_1	b_2
1	-1.75	1.35	0.513	0.366
2	-4.528	8.883	9.365	0.249
3	13.11	63.53	108.2	0.198
4	96.69	410.0	690.3	0.201
5	2460	1375	4072	0.151
6	11929	4356	16614	0.130
7	50756	6105	54849	0.0750
8	103103	10566	109600	0.0675
9	201232	10089	205085	0.0399
10	291025	8894.1	293206	0.0265
11	370302	8786	372025	0.0212
12	449537	8740	450961	0.0177
13	529059	7804	530056	0.0137

Table 1: Optimal values of the coefficients in the linear and exponential models

3.

The values of the functions F_{lin} and F_{exp} at the optimal values for $\mathbf{a}$ resp. $\mathbf{b}$ can be compared with each other. For the values $\mathbf{a}, \mathbf{b}$ given in Table 1, we have $F_k^{lin}(\mathbf{a}) > F_k^{exp}(\mathbf{b})$ for $1 \leq k \leq 6$ and $F_k^{lin}(\mathbf{a}) < F_k^{exp}(\mathbf{b})$ for $6 \leq k \leq 13$.

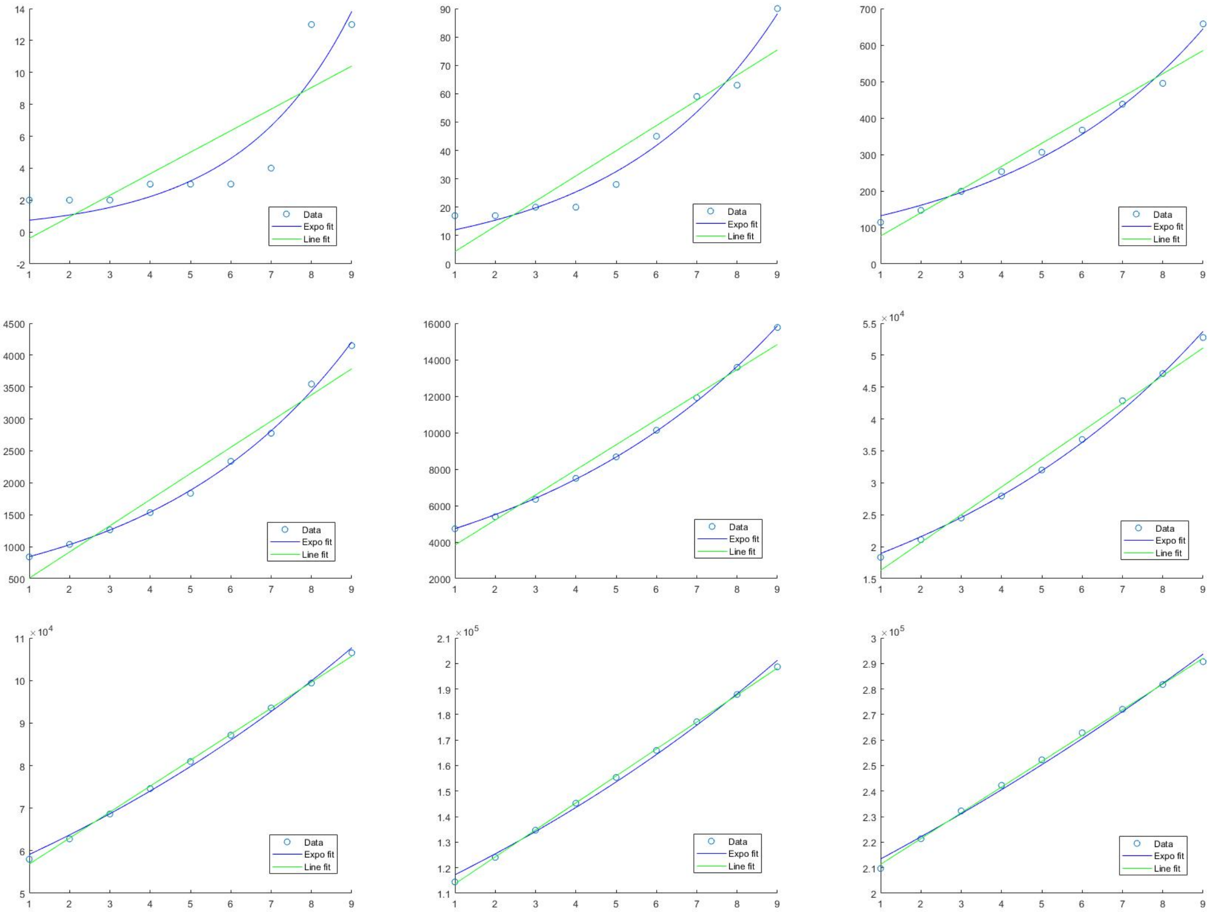

Figure 3: Graphs illustrating the behaviour at the start of the pandemic in Russia as well as the transition from exponential to linear growth.

3 Refinements

The transition from exponential growth to linear growth can be called *straightening* the curve. As we have seen in the previous section for Russia the growth seems to stay linear. We admit that the picture may depend on the choice of the *basic period*, which is the number of data point used for fitting the functions. When we choose the basic period to be 10 days instead of 9, we find the first four periods to be exponential, the fifth linear, the sixth exponential, and the subsequent periods are linear.

The result may also depend on which day you start each basic period. Therefore one should start a new period every day. To determine the phase the pandemic is in at a particular day k, we take into account data from the period from day $k - \lfloor p/2 \rfloor$ till day $k - \lfloor p/2 \rfloor - (p+1 \mod 2)$, where p is the number of days in one basic period.

It can be seen that for countries such as Australia, see Figure 2c, the data bends down. This is known as *flattening* the curve. A mathematical model that describes such a transition is the logarithmic model

$$y = f_{\mathbf{c}}(x) = c_1 + c_2 \log(x).$$

We propose to compare the values of the objective functions

$$F_{k,p}^{lin}(\mathbf{a}) = \sum_{i=-\lfloor p/2 \rfloor}^{\lfloor p/2 \rfloor - (p+1 \mod 2)} \left(y_{k+i} - f_{\mathbf{a}}(x_{k+i})\right)^2,$$

$$F_{k,p}^{exp}(\mathbf{b}) = \sum_{i=-\lfloor p/2 \rfloor}^{\lfloor p/2 \rfloor - (p+1 \mod 2)} \left(y_{k+i} - f_{\mathbf{b}}(x_{k+i})\right)^2,$$

and

$$F_{k,p}^{log}(\mathbf{c}) = \sum_{i=-\lfloor p/2 \rfloor}^{\lfloor p/2 \rfloor-(p+1 \mod 2)} (y_{k+i} - f_{\mathbf{c}}(x_{k+i}))^2$$

at their respective minimisers, to find out in what phase the data is at day k, where $\lfloor p/2 \rfloor < k \leq n - \lfloor p/2 \rfloor - (p+1 \mod 2)$ and n is the total number of data points. For COVID-19 confirmed cases in NSW data the results, with $p = 7$ are colour coded in Figure 4.

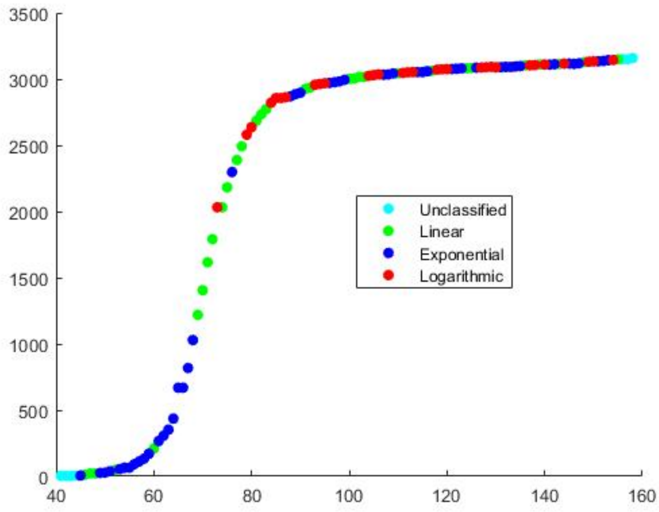

Figure 4: Different phases for COVID-19 confirmed cases in NSW

In Figure 4 an exponential period can be clearly identified, as well as a linear period. The method seems to be too sensitive to properly classify the (rather flat) tail of the curve. Let us consider three consecutive days at the tip of the tail, see Figure 5.

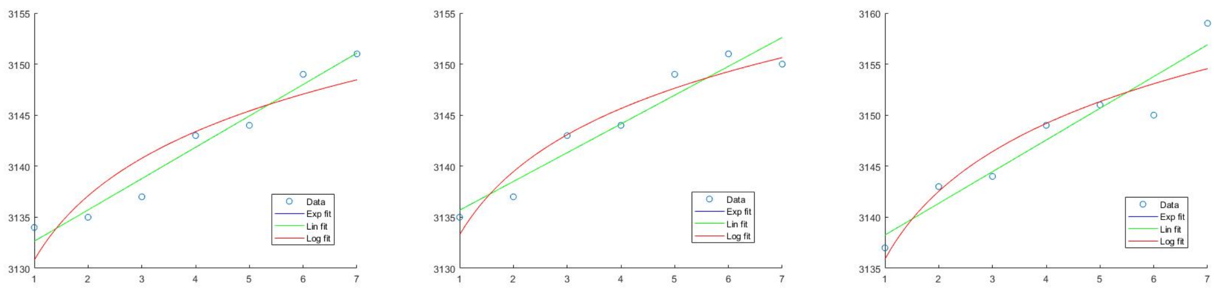

Figure 5: Fitted curves for 3 consecutive days, with 3 different phases.

The data in Figure 5 is classified, from left to right, as exponential, logarithmic, and linear. The first thing to note is that the eye is not able to distinguish between the exponential curves and the linear ones. The reason for this is that the exponent is too small. Indeed, using a Taylor series:

$$b_1 \exp(b_2 x) = b_1(1 + b_2 x + \frac{b_2^2}{2} x^2) \approx b_1 + b_1 b_2 x$$

for all $0 < x \leq p$, if $b_2 << \frac{2}{p}$. The linear and exponential fits are quite different from the logarithmic ones. To allow for natural variation in the data, it makes sense to only classify the growth as logarithmic if the minimal value of F^{log} is significantly lower than the minimal value of F^{lin}. Requiring the ratio of minimal values F^{log}/F^{lin} to be smaller that 0.9 (as well as $F^{exp}/F^{lin} < 0.9$) yields the classifications of growth phases for NSW, Russia and Italy in Figure 6.

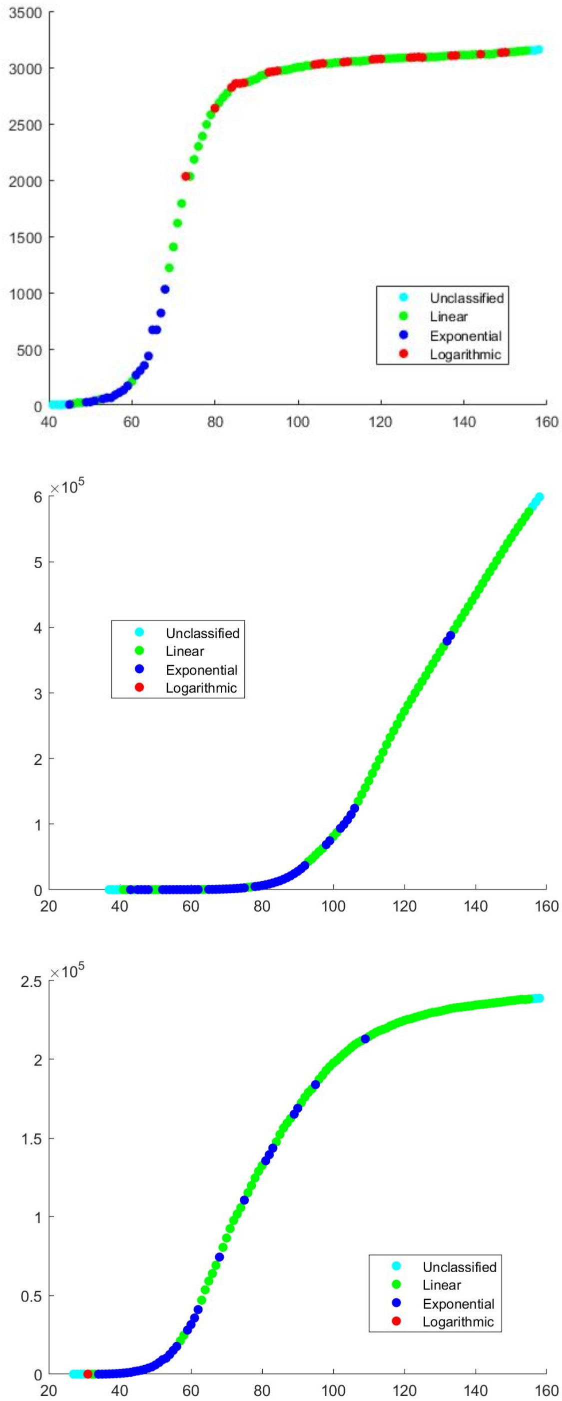

Figure 6: Classification of growth phases for NSW, Russia, and Italy ($p = 7$).

The exponential periods can be read of from these graphs.

NSW: 27 days (from day 42 till day 68),
Russia: 64 days (from day 43 till day 106),
Italy: 29 days (from day 34 till day 62).

As a final remark: flattening of the curve does not only take place in the logarithmic phases. Looking at the values of the coefficient b_2 in Table 1 they are decreasing from period 4 onwards, the values of coefficient a_2 is decreasing from period 8 onwards.

References

[1] A. Close, OPT Project, report for MAT5OPT, La Trobe University, 11pp.

[2] https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases