Handling discrete citation data and Kao's Nobel Prize winning article in fiber optics communication

Yuxian Liu [1,2] and Ronald Rousseau [2,3,4]

[1]Periodical Division, Reading Section, Library of Tongji University, Siping Street 1239, 200092 Shanghai, PR China
E-mail: yxliu@lib.tongji.edu.cn

[2]University of Antwerp, IBW, Venusstraat 35, B-2000 Antwerpen, Belgium

[3] KHBO (Association K.U.Leuven), Industrial Sciences and Technology, Zeedijk 101, B-8400 Oostende, Belgium

[4] K.U.Leuven, Dept. Mathematics,
Celestijnenlaan 200B, B-3000 Leuven (Heverlee), Belgium
E-mail: ronald.rousseau@khbo.be

Abstract

We discuss technical problems in handling discrete data such as citation data. Using Kao's citation data it is shown that splines can provide an everywhere differentiable description over the whole data interval. We find that citations to Kao's Nobel Prize winning article follow the main developments of the field of fiber optics communication.

Introduction

Charles Kuen Kao (Gao Kun in pinyin) received the Nobel Price (actually one half) in Physics 2009 for his contribution to the development of low-loss optic fiber used in optical fiber communication systems (Class for Physics, 2009). Although Kao has published more than 100 articles in his career, we were able to find only 32 of them in the Web of Science (WoS). In this contribution we study citations to his most important article:

> *Dielectric-fibre surface waveguides for optical frequencies*
> *Kao, K.C and Hockham, G.A.*
> *Proceedings of the Institution of Electrical Engineers – London, (1966), 113(7), 1151-1158.*

This article has been republished as:

Dielectric-fiber surface waveguides for optical frequencies
Kao, K.C and Hockham, G.A.
IEE Proceedings – J Optoelectronics, (1986), 133(3), 191-198.

These two articles will be referred to as the Kao-Hockham article. Together they received 199 citations (197+2) during the period [1966-2008]. These are the citations we study. Compared to most Nobel laureates this number is rather small (Kao's WoS-based *h*-index by the end of 2009 was 5), confirming the fact that citations in engineering are often quite low (Podlubny, 2005).

Data

We collected all citations received by the Kao-Hockham article (WoS data). These are shown in Table 1.

Table 1. Citations to Kao-Hockham: time distribution

Year	Age since publication	Cumulative citations, C
1966	0	1
1967	1	1
1968	2	3
1969	3	4
1970	4	13
1971	5	18
1972	6	22
1973	7	31
1974	8	37
1975	9	44
1976	10	56
1977	11	61
1978	12	65
1979	13	72
1980	14	77
1981	15	83
1982	16	90
1983	17	94
1984	18	100
1985	19	109
1986	20	120
1987	21	123
1988	22	125
1989	23	127

1990	24	132
1991	25	133
1992	26	135
1993	27	140
1994	28	140
1995	29	143
1996	30	146
1997	31	147
1998	32	153
1999	33	154
2000	34	163
2001	35	165
2002	36	170
2003	37	175
2004	38	179
2005	39	184
2006	40	190
2007	41	194
2008	42	199

Problem statement

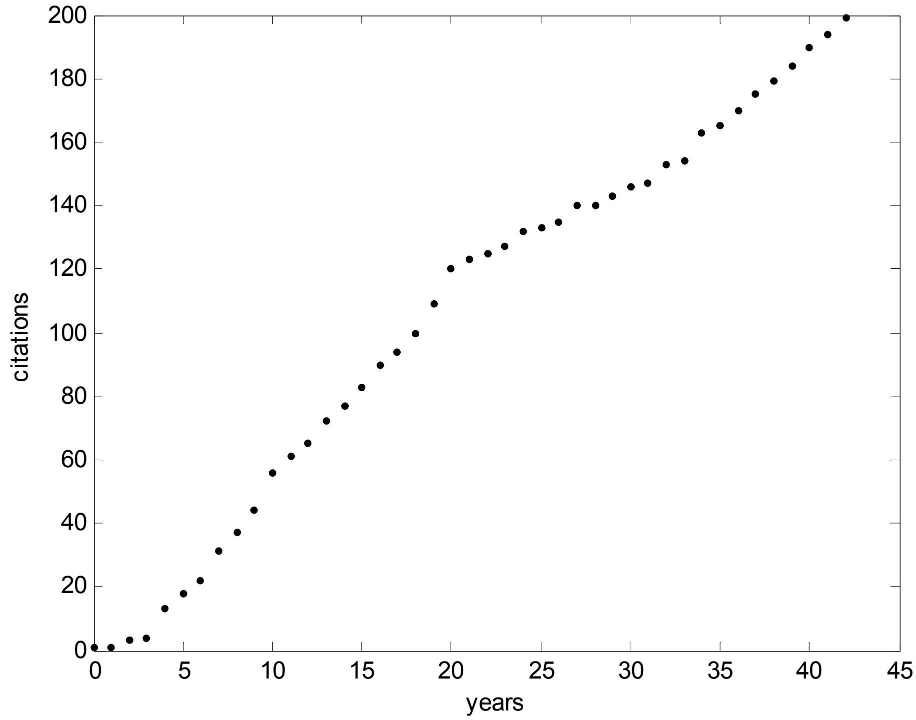
We are interested in the study of citations in terms of change (velocity) and change of change (acceleration) over time. Note that the received number of citations per period is just the change in the total number of citations, hence the change in this number during this period. Change per unit of time can be considered a velocity. This explains the terminology we use in this contribution.

For differentiable functions, F, velocity is obtained as the derivative of F, denoted as F', and the acceleration as the derivative of F', denoted as F". However, in citation analysis it is customary (often by necessity) to collect data on a yearly basis. Even if data are collected more frequently the result is always a discrete sequence. This leads to the problem of finding an everywhere differentiable function that describes given discrete data points. A well-known solution is to determine a best-fitting exponential, power (including a linear function) or logistic function (MacRae, 1969; Cano & Lind, 1991; Egghe & Rao, 1992). These may all be acceptable models for a cumulative citation function over a certain interval. Such approaches, however, assume that all information that is available in the data resides in the fitted curve. As we are interested in local changes in the curve we take another approach.

Forward and backward differences are used in numerical analysis techniques for finding interpolating polynomials. An interpolating polynomial for a set of n points $(x_j,f_j)_{j = 1, \ldots n}$ is a polynomial P(x) such that $P(x_j) = f_j$, for j =1,… ,n. It is, moreover well known that there exists exactly one interpolating polynomial of degree n-1 or lower for a given set of points $(x_j,f_j)_{j = 1, \ldots n}$ (Hildebrand, 1956; Rousseau, 1997). Usually, however this procedure leads to polynomials of high degree showing many fluctuations unrelated to the data set one wants to describe. For our purposes, interpolating cubic splines seem to be a better approach. We recall that interpolating cubic splines are 1) interpolating functions (they pass through the given data points), 2) cubic, i.e. polynomials of third degree, and 3) have first and second derivatives in the data points. This means that they connect data points in a smooth way. They, moreover, satisfy a best-approximation property which tends to reduce the curvature of interpolating cubic splines (Greville, 1967; Nürnberger, 1989, Rousseau, 1997). Because of these properties it is meaningful to calculate derivatives in the data points and even in between. We recall that the equations describing such splines differ between data points (see appendix for examples).

Results

Equations for the resulting interpolating splines for the cumulative citation curve of the Kao-Hockham article are given in the appendix. The obtained curve of cubic spline interpolations together with the original data are shown in Fig 1.



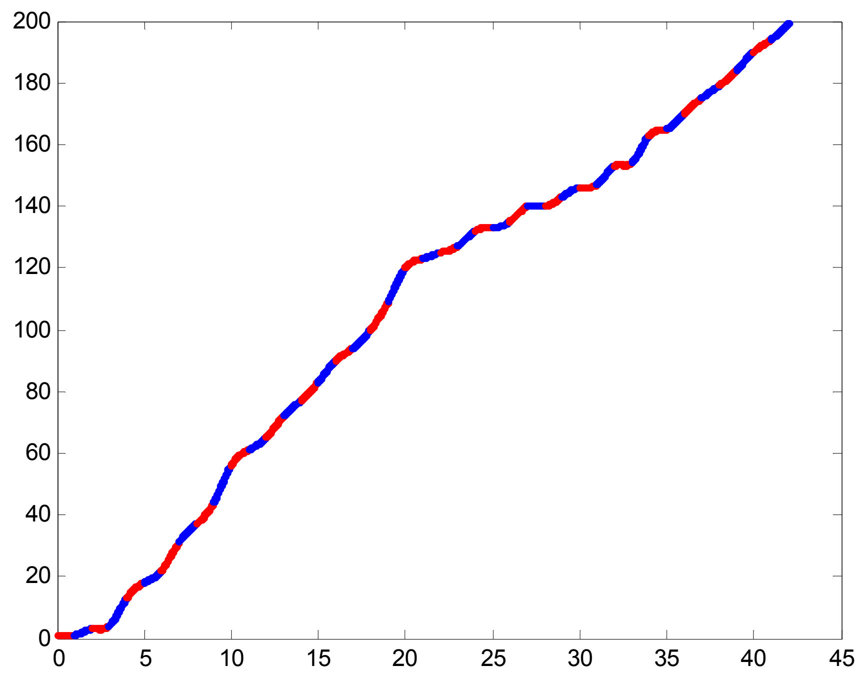


Fig. 1 Kao’s cumulative citation data and spline interpolations

We next calculate the first and second order derivative functions of these splines, shown in Figs. 2 and 3.

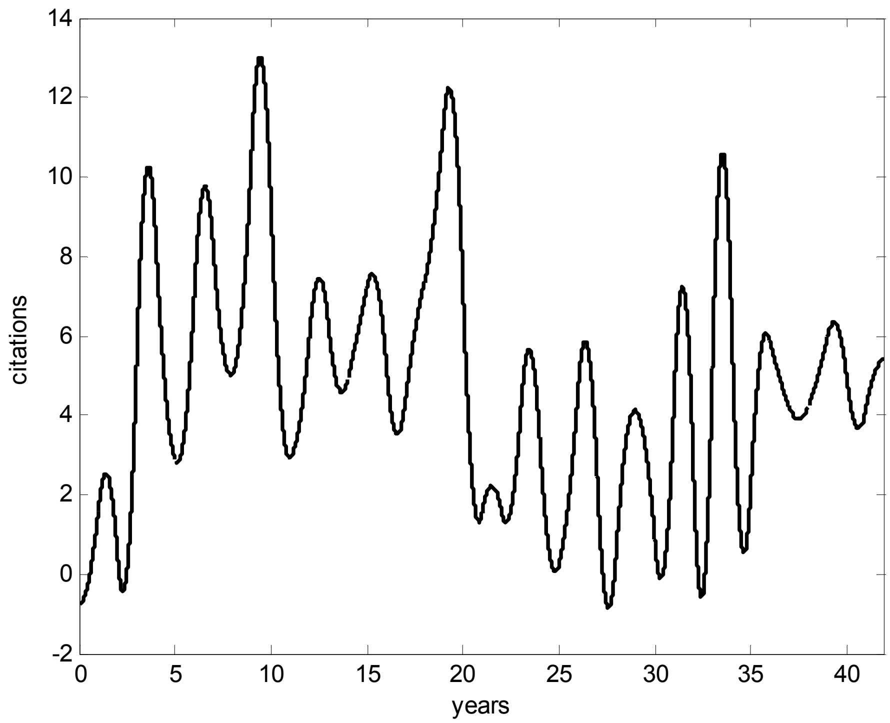


Figure 2. First order derivative functions of interpolating splines

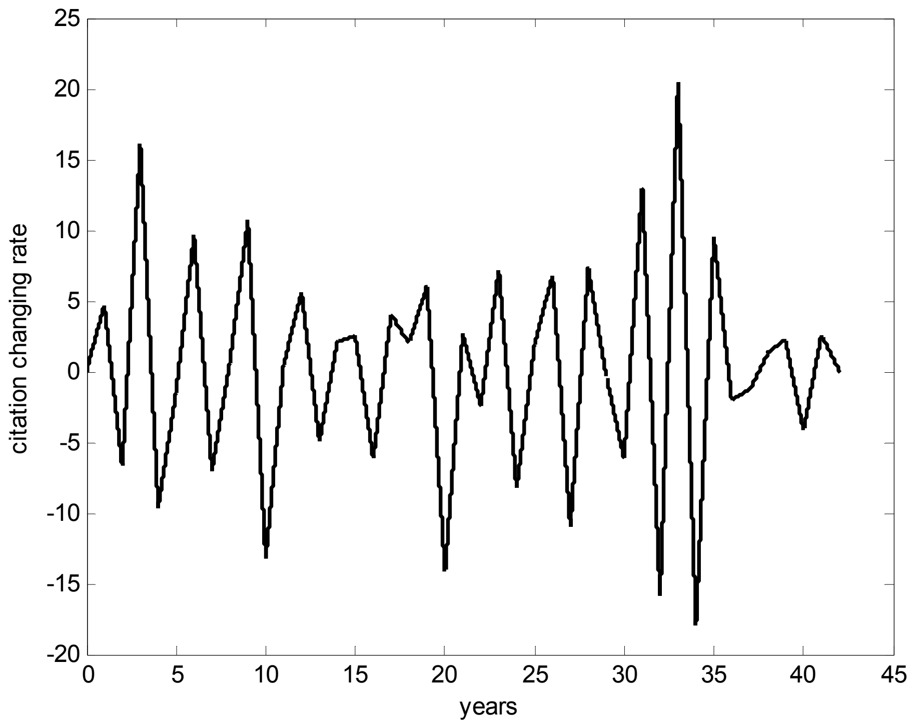


Figure 3. Second order derivative function of interpolating splines

Analysis

Relation between citation velocity and acceleration

We now discuss the relation between velocity and acceleration based on the interpolating splines (see Fig. 4).

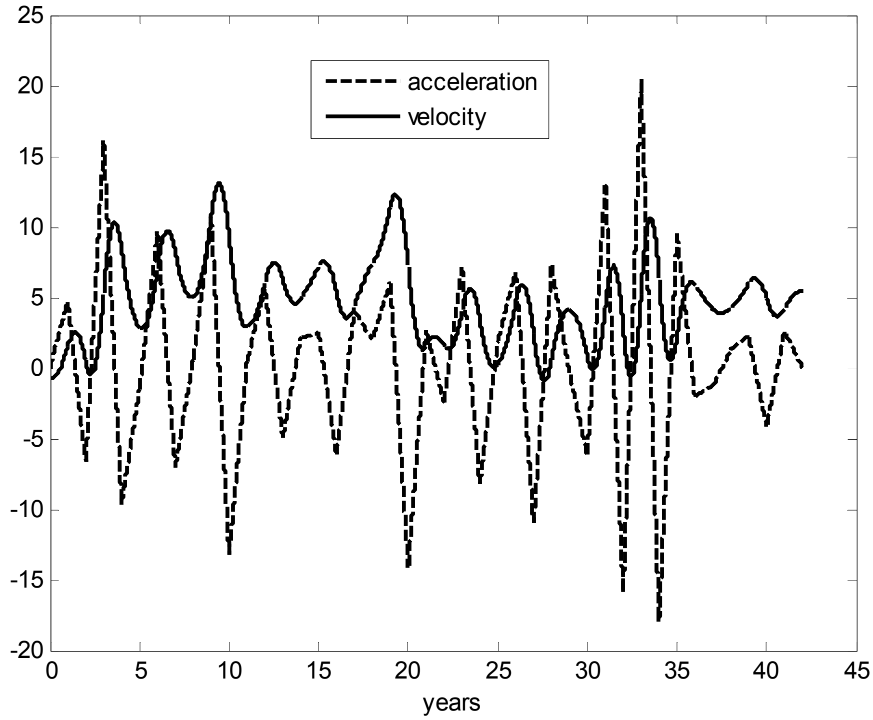


Figure 4. Velocity and acceleration based on spline data

The velocity curve goes up and down, somewhat like a sine function; the accelation curve also goes up and down, but with a difference of one step, somewhat like a cosine function. Out of curiosity we shifted the acceleration curve over one unit to see if data become more synchronised in this way. This was indeed the case, leading to a Pearson correlation coefficient that went from -0.00033 to 0.42. Naturally acceleration peaks precede velocity peaks: positive acceleration leads to a higher velocity; negative acceleration (deceleration) leads to lower velocity. When studying the cumulative citation function of Kao's article we did an interesting observation, related to the development of the field of fiber-optics communication. For this reason we first give a short overview of the history of the field.

Fiber optics communication

In (Gambling, 2000) and in Wikipedia (2010) we find, among other information, a short history of the field. It is roughly divided into five periods, corresponding to early developments and four generations of fiber-optic communication systems. The first period begins with the Kao-Hockham article (1966) leading to the first development of optical fiber by Corning Glass Works in 1970. In the second period, starting from 1975, the first generation of commercial fiber-optic communications system was developed. In the third period, starting in early 1980 the second generation of fiber-optic communication for commercial use was developed. Then we come to the third- and fourth generation fiber-optic systems, causing a real revolution in the increase of data capacity. During the period 1985-1987 David Payne and his group at Southampton (Mears et al., 1987) invented the erbium fiber amplifier (EFA) making it possible, among other things, to install TransAtlantic and TransPacific underwater cables with a capacity equivalent of 600,000 telephone circuits at very low prices (Gambling, 2000). (By the way, Mears' article has been cited more than Kao's, illustrating the importance of this finding). From 1992 to 2001 system capacity doubled every six months, leading around 2000-2001 to an oversupply in capacity, referred to as the so-called fiber glut. However, it turned out that people need more and more capacity and fiber optical communication systems are nowadays developing steadily.

Back to the cumulative citation curve

Figure 5 shows the spline-fitted cumulative citation curve, together with a best-fitting linear line. This straight line intersects the citation curve in two points: once near 1975 and once near the year 2000, two important dates in the field's history. The year 1975 corresponds to the transition from academic research to research on real systems, while the year 2000 corresponds to the year in which economic limitations slowed down practical progress. In the first period, 1966-1975 the citation curve is roughly convex (positive second derivative is positive) showing accelerated growth; in the second period, 1975-2000, there is first a linear growth (faster than that of the regression line), followed by a concave part (second derivative is negative). This period corresponds to a technical period (not an academic one). Application periods are indeed often characterized by a slowing down of academic publications (Small & Upham, 2007). Then during the last part the citation curve is roughly parallel to the regression line. This last part corresponds to normal commercial operation of fiber optic communication. Moreover, local extrema occur in 1970 and 1986, two other important data in the history of fiber optics. Indeed 1970 is the year in which the first commercial optical fiber was realized by Corning Glass Works and 1986 corresponds to the invention of EFA. One may say that breakthroughs in fiber optic communication promote the diffusion of Kao's ideas, while technological and economic difficulties reduce its diffusion. Consequently, Kao's cumulative citation curve reflects the history of its field.

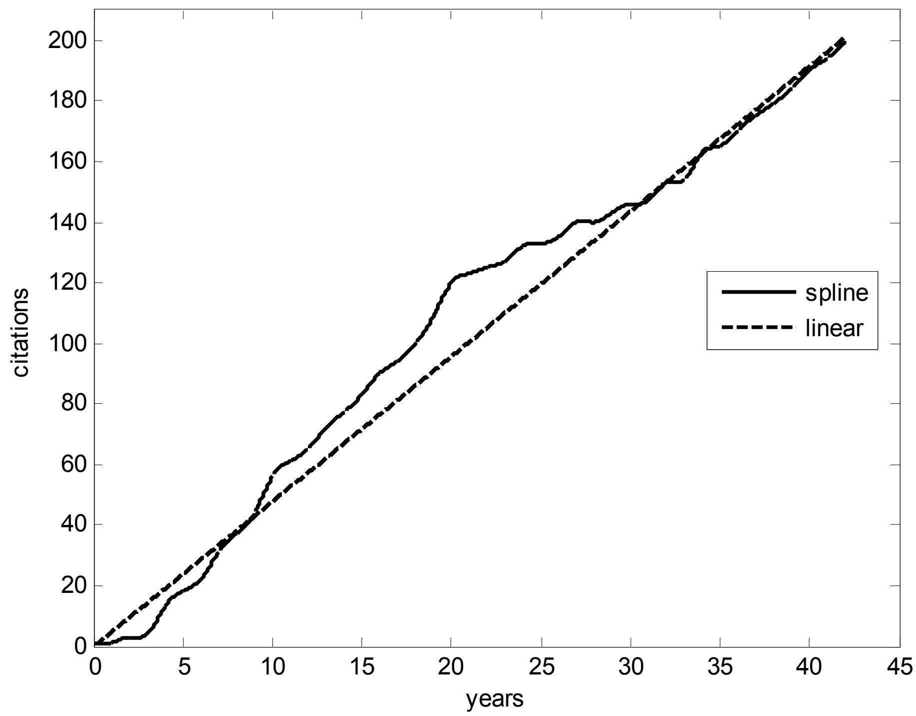


Fig. 5 Spline-fitted cumulative citation curve, best-fitting linear curve and indications of important years

Conclusion

Using Kao's citation data as a case study we discussed technical problems in handling discrete data. It is shown that splines can provide an everywhere differentiable description of discrete data. As an unexpected bonus we observed that citations to Kao's Nobel Prize winning article follow the main developments of the field of fiber optics communication.

Acknowledgement

The authors thank Raf Guns (UA, Belgium) for helpful comments.

References

Cano, V. & Lind, N.C. (1991). Citation life-cycles of ten citation-classics. Scientometrics, 22(2), 297-312.

Class for Physics of the Royal Swedish Academy of Sciences (2009). Two revolutionary optical technologies. Royal Swedish Academy of Sciences. Available at:
nobelprize.org/nobel_prizes/physics/laureates/2009/phyadv09.pdf

Egghe L. & Rao, I.K.R. (1992). Classification of growth models based on growth rates and its applications. Scientometrics, 25(1), 5-46.

Gambling, W.A. (2000). The rise and rise of optical fibers. IEEE Journal of Selected Topics in Quantum Electronics, 6(6), 1084-1093.

Greville, T.N.E. (1967). Spline functions, interpolation, and numerical quadrature. In: Mathematical methods for digital computers, vol. II (A. Ralston & H.S. Wilf, eds). New York:Wiley, 156-168.

Hildebrand, F.B. (1956). Introduction to numerical analysis. New York: McGraw-Hill.

MacRae, Jr. D. (1969). Growth and decay curves in scientific citations. American Sociological Review, 34(5), 631-635.

Mears, R.J., Reekie, L., Jauncey, I.M. and Payne, D.N. (1987). Low-noise erbium-doped fiber amplifier operating at 1.54 μm. Electronics Letters, 23(19), 1026-1028.

Nürnberger, G. (1989). Approximations by spline functions. Berlin: Springer-Verlag.

Podlubny, I. (2005). Comparison of scientific impact expressed by the number of citations in different fields of science. Scientometrics, 64(1), 95-99.

Rousseau, R. (1997). Numerical mathematics (in Dutch). Course notes, Oostende: KHBO.

Small, H. & Upham, P. (2007). Citation structure of an emerging research area: organic thin film transistors. Proceedings of ISSI 2007 (Torres-Salinas & Moed, eds.). Madrid: CINDOC-CSIC, p. 718-725.

Wikipedia (2010). http://en.wikipedia.org/wiki/Fiber-optic_communication (text last modified on 15 March 2010).

Appendix.

$$
s(x)=\begin{cases}
0.78x^3-0.78x+1 & x\in[0,\ 1]\\
-1.89x^3+8.02x^2-8.80x+3.67 & x\in[1,\ 2]\\
3.80x^3-26.12x^2+59.47x-41.84 & x\in[2,\ 3]\\
-4.29x^3+46.64x^2-158.79x+176.42 & x\in[3,\ 4]\\
1.36x^3-21.12x^2+112.24x-184.95 & x\in[4,\ 5]\\
1.86x^3-28.59x^2+149.60x-247.23 & x\in[5,\ 6]\\
-2.78x^3+54.92x^2-351.45x+754.88 & x\in[6,\ 7]\\
1.28x^3-30.33x^2+245.27x-637.46 & x\in[7,\ 8]\\
1.68x^3-39.98x^2+322.48x-843.36 & x\in[8,\ 9]\\
-3.99x^3+113.05x^2-1054.81x+3288.52 & x\in[9,\ 10]\\
2.28x^3-75.03x^2+825.99x-2980.83 & x\in[10,\ 11]\\
0.87x^3-28.51x^2+314.33x-1104.75 & x\in[11,\ 12]\\
-1.76x^3+66.21x^2-822.36x+3442.04 & x\in[12,\ 13]\\
1.17x^3-48.24x^2+665.43x-3005.05 & x\in[13,\ 14]\\
0.07x^3-1.75x^2+14.66x+31.85 & x\in[14,\ 15]\\
-1.44x^3+66.09x^2-1002.92x+5119.74 & x\in[15,\ 16]\\
1.70x^3-84.48x^2+1406.17x-7728.73 & x\in[16,\ 17]\\
-0.34x^3+19.55x^2-362.35x+2292.89 & x\in[17,\ 18]\\
0.68x^3-35.66x^2+631.46x-3669.97 & x\in[18,\ 19]\\
-3.37x^3+195.18x^2-3754.62x+24108.56 & x\in[19,\ 20]\\
2.81x^3-175.45x^2+3658.16x-25309.96 & x\in[20,\ 21]\\
-0.85x^3+55.08x^2-1183.12x+8578.99 & x\in[21,\ 22]\\
1.61x^3-107.28x^2+2388.74x-17614.65 & x\in[22,\ 23]\\
-2.57x^3+181.18x^2-4245.74x+33249.68 & x\in[23,\ 24]\\
1.69x^3-125.69x^2+3119.17x-25669.64 & x\in[24,\ 25]\\
0.82x^3-60.60x^2+1491.86x-12108.72 & x\in[25,\ 26]\\
-2.97x^3+235.06x^2-6195.29x+54513.25 & x\in[26,\ 27]\\
3.06x^3-253.36x^2+6992.17x-64173.89 & x\in[27,\ 28]\\
-1.27x^3+110.29x^2-3190.15x+30861.10 & x\in[28,\ 29]\\
-0.98x^3+85.33x^2-2466.20x+23862.95 & x\in[29,\ 30]\\
3.20x^3-291.00x^2+8823.75x-89036.56 & x\in[30,\ 31]\\
-4.81x^3+454.19x^2-14277.13x+149672.53 & x\in[31,\ 32]\\
6.06x^3-589.31x^2+19114.60x-206505.95 & x\in[32,\ 33]\\
-6.41x^3+644.83x^2-21611.90x+241485.62 & x\in[33,\ 34]\\
4.58x^3-476.60x^2+16516.67x-190638.23 & x\in[34,\ 35]
\end{cases}
$$

$$s(x)=\begin{cases} -1.93x^3+207.12x^2-7413.35x+88545.33 & x\in[35,36] \\ 0.12x^3-14.49x^2+564.36x-7187.22 & x\in[36,37] \\ 0.43x^3-48.24x^2+1813.09x-22588.13 & x\in[37,38] \\ 0.16x^3-17.68x^2+651.81x-7878.59 & x\in[38,39] \\ -1.07x^3+126.44x^2-4968.74x+65188.60 & x\in[39,40] \\ 1.12x^3-136.98x^2+5567.92x-75300.22 & x\in[40,41] \\ -0.42x^3+53.52x^2-2242.60x+31443.55 & x\in[41,42] \end{cases}$$