

EEN UNIFORM ALGORITME VOOR DISCRETE SMOOTHING  
EN INTERPOLATIE

P.H.C. Eilers\*

Samenvatting

Een door Whittaker voorgesteld algoritme voor discrete smoothing kan met enige ingrepen gebruikt worden als een uniform gereedschap voor smoothing en interpolatie. In het oorspronkelijke algoritme kunnen ontbrekende waarden verwerkt worden door middel van gewichten. Interpolatie kan als een probleem met zeer veel ontbrekende waarden en geringe smoothing geïnterpreteerd worden. De basis van de smoothing is een kleinste-kwadraten functie, waarin de  $k$ -de differentie van de oplossing opgenomen is. Door de differenties te wegen kan men knikken of sprongen in een reeks aanvatten. Een gewogen gemiddelde van eerste en tweede differenties maakt een vloeiende overgang van kubische, via exponentiële naar lineaire interpolatie mogelijk.

\* Dienst Centraal Milieubeheer Rijnmond  
's-Gravelandseweg 565  
3119 XT Schiedam  
(010) 42 73 699

## 1. Inleiding

Zeer veel gegevens komen beschikbaar in een vorm die ordening naar een of andere parameter zoals tijd of afstand zinvol maken. Het is zelden het geval dat een grafische weergave van de ruwe geordende gegevens bevredigt. Meestal is het resultaat onplezierig "rafelig" van aanzien door de aanwezigheid van ruis en/of fluctuaties. Ook is men vaak geïnteresseerd in afgeleide zaken, zoals een trendlijn, geïnterpoleerde waarden, eerste of tweede afgeleide. Om deze vragen te kunnen beantwoorden is een of ander vorm van "smoothen" nodig. Er bestaan zeer veel technieken om op discrete onderlinge afstanden gegeven waarden te smoothen. Het doel van dit artikel is aan te geven dat één daarvan, discrete smoothing, een bijzonder flexibel en effectief algoritme oplevert, dat in vele situaties zijn nut kan bewijzen. Discrete smoothing is oorspronkelijk voorgesteld door Whittaker (1923) voor het aanpassen van getallen in sterftetabellen. Het algoritme kan eenvoudig aangepast worden aan het ontbreken van waarnemingen. Men kan dit zo ver doorvoeren dat de meeste waarnemingen "ontbreken". Dan blijkt het algoritme als een effectieve interpolator te werken. De problemen die men oplost met smoothing splines (Reinsch, 1967, 1971) of exponentiële splines (Rentrop, 1980) kan men ook aanpakken met discrete smoothing. Ook voor knikken en sprongen in de datareeks kan men eenvoudig voorzieningen treffen.

## 2. Discrete smoothing

Stel dat een equidistant bemonsterde reeks  $y_i$ ,  $i = 1..m$ , gegeven is, zonder ontbrekende waarden. De lokale mate van verandering (de helling) van de reeks wordt gegeven door de eerste differentie (Fröberg, 1969):

$$\Delta y_i = y_i - y_{i-1} \quad i = 2..m \quad (2.1)$$

De lokale kromming kan uitgedrukt worden door een niveau dieper te gaan:

$$\Delta^2 y_i = y_i - 2y_{i-1} + y_{i-2} \quad i = 3..m \quad (2.2)$$

We kunnen ook centrale dubbele differenties gebruiken:

$$\delta^2 y_i = y_{i-1} - 2y_i + y_{i+1} \quad i = 2..m-1 \quad (2.3)$$



Globale maten voor de veranderingen in de reeks zijn:

$$T_1 = \sum_{i=2}^m (\Delta y_1)^2 \quad \text{en} \quad T_2 = \sum_{i=2}^{m-1} (\delta^2 y_1)^2 \quad (2.4)$$

Een glad verlopende reeks wordt gekenmerkt door een lage waarde van  $T_2$  en, in mindere mate,  $T_1$ . Het voorbeeld van een rechte lijn is illustratief. Daar geldt  $\Delta y_1 = c$ , met  $c$  een konstante. Dan is  $\delta^2 y_1$  dus identiek nul. Een horizontale rechte lijn geeft dus  $T_1 = 0$ , iedere rechte lijn geeft  $T_2 = 0$ .

Al zeer lang geleden stelde Whittaker (1923) vóór de "graduation" - wat we tegenwoordig "smoothing" noemen - van een reeks te baseren op het mede minimaliseren van termen als  $T_1$  en  $T_2$ . Cleveland (1983) verwijst naar Henderson (1924) voor een identiek voorstel. Deze referentie blijkt echter moeilijk te bemachtigen. Het smoothing-probleem wordt geformuleerd als: minimaliseer

$$S = \sum_{i=1}^m (y_1 - z_1)^2 + \beta \sum_{i=k+1}^m (\Delta^k y_1)^2 \quad (2.5)$$

De betekenis van de termen van  $S$  is duidelijk: de eerste term meet hoe goed de reeks  $z$  de reeks  $y$  benadert, terwijl de tweede term meet hoe glad de reeks  $z$  verloopt. De balans tussen beide termen wordt geregeld door de parameter  $\beta$ . Hoe groter  $\beta$ , des te gladder verloopt de reeks  $z$ , maar des te meer zal ze afwijken van de reeks  $y$ .

Whittaker leidt een aantal eigenschappen af van de oplossing van (2.5), waarvan de bruikbaarste zijn dat  $\sum y_1^j = \sum z_1^j$  als  $j < k$ .

Het voorstel van Whittaker, met  $k=2$ , wordt nu uitgebreid met gewichten in beide termen van  $S$ . De doelfunctie wordt dan:

$$S = \sum_{i=1}^m w_1 (y_1 - z_1)^2 + \beta \sum_{i=2}^{m-1} v_1 (\delta^2 z_1)^2 \quad (2.6)$$

Het doel van de gewichten  $w_1$  is, de mate van aanpassing lokaal te beïnvloeden. Dat vereist voorkennis van de spreiding in  $y$ .

Een extreem, maar vaak voorkomend, geval is het ontbreken van waarnemingen. Dan kan  $y_i$  op iedere gewenste waarde gezet worden als  $w_i$  nul is, zodat alleen de tweede term van  $S$  telt.

Het doel van de gewichten  $v_i$  is, de toegestane kromming lokaal te beïnvloeden. De voornaamste toepassingen liggen ook hier bij extreme maar praktisch relevante situaties: een knik of een sprong in de gegeven reeks.

De waarden  $z_i$  die  $S$  minimaal maken, volgen uit het nul stellen van de partiële afgeleiden. Het resultaat is:

$$a_i' z_{i-2} - b_i' z_{i-1} + (\alpha w_i + c_i) z_i - b_i z_{i+1} + a_i z_{i+2} = \alpha w_i y_i \quad i = 1..m \quad (2.7)$$

met:

$$a_i' = v_{i-1} ; \quad b_i' = 2 (v_{i-1} + v_i) ; \quad \alpha = 1/\beta \quad (2.8)$$

$$a_i = v_{i+1} ; \quad b_i = 2 (v_i + v_{i+1}) ; \quad c_i = v_{i-1} + 4 v_i + v_{i+1}$$

en:

$$v_0 = v_1 = v_m = v_{m+1} = 0 \quad (2.9)$$

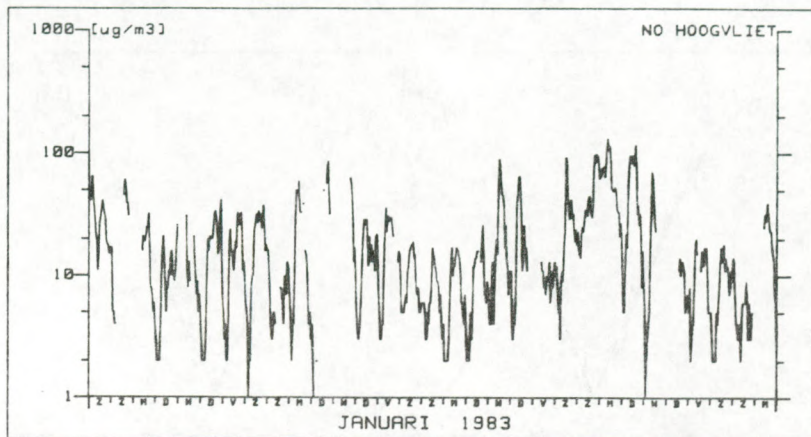
De matrix van het stelsel lineaire vergelijkingen (2.7) heeft een bandstructuur; de hoofddiagonaal en aan weerszijden twee neven-diagonalen zijn niet nul. Met een variant op de Cholesky-decompositie, de  $U^T DU$ -decompositie (Bierman, 1977), is het stelsel eenvoudig op te lossen.

### 3. Toepassingen van smoothing

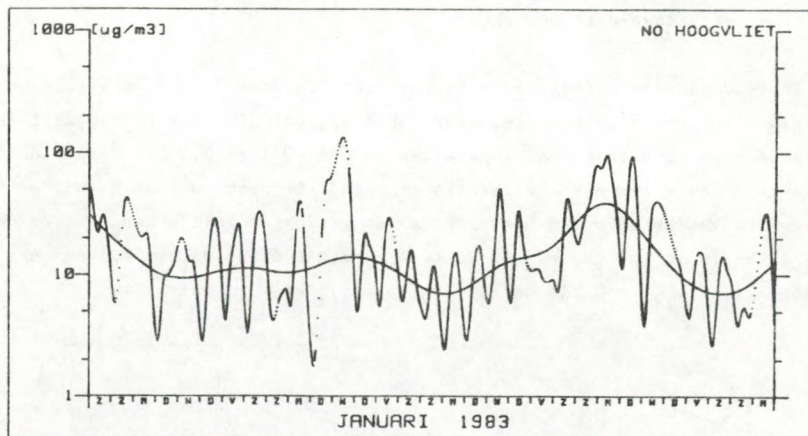
Figuur 3.1 toont het verloop van stikstofmonoxide in Hoogvliet in januari 1983, zoals gemeten door de DCMR. Het betreft uurgemiddelde concentraties. De verticale schaal is logaritmisch; smoothing vindt plaats met de logaritmen van de concentraties. Figuur 3.2 toont de gesmoothede reeks, voor  $\beta = 100$  en  $\beta = 10^6$ . Waar in de originele reeks punten ontbraken vindt interpolatie plaats. Daar zijn de lijnen gestippeld (voor  $\beta = 10^6$  liggen de punten zo dicht bijeen dat ze samenvloeien).

Uit de figuur blijkt dat lichte smoothing het dagelijks ritme (het verkeer is een belangrijke bron van stikstofoxiden) accentueert en dat sterke smoothing de trend oplevert.



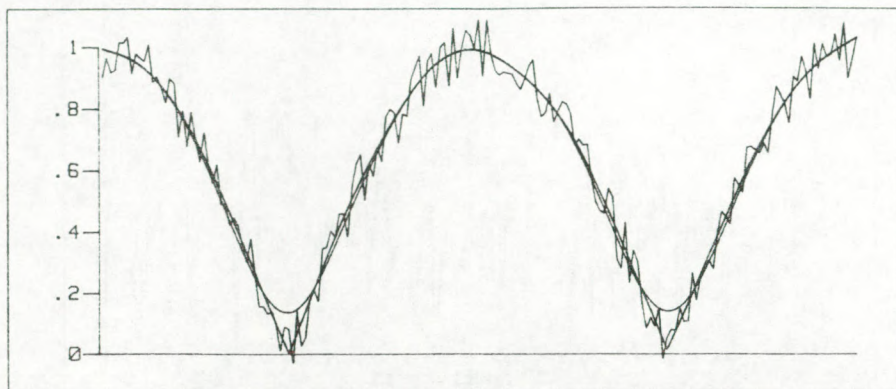


Figuur 3.1: Uurgemiddelde concentraties van stikstofoxiden.



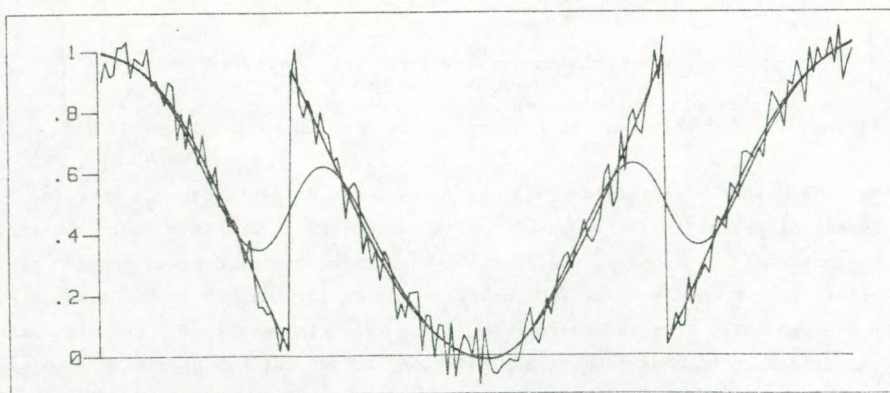
Figuur 3.2: Smoothing van stikstofoxiden met  $\beta = 100$  en  $\beta = 10^6$ .

Nu volgen enkele voorbeelden van het gebruik van de gewichten  $v_1$ , voor de tweede differenties. In figuur 3.3 zijn gesimuleerde data ( $m = 180$ ) getekend, bestaande uit de absolute waarde van een sinusfunctie plus pseudo-random getallen tussen  $-0.1$  en  $0.1$ . Voor de reeks met de afrondingen in de dalen is  $v_1 = 1$  gesteld, voor de andere reeks is op twee plaatsen  $v_1 = 0$  gesteld. Daar treedt een knik in de gesmoothede reeks op, zonder dat het glatte verloop aan weerszijden van de knik verstoord wordt.



Figuur 3.3: Gesimuleerde reeks met knikken en gesmoothede reeksen met en zonder aangepaste gewichten.

Een voorbeeld van een sprong geeft figuur 3.4. Ook deze data zijn gesimuleerd. Als basis dient een sinusfunctie, waar 1 bij opgeteld is waar ze negatief is. Vervolgens zijn er pseudo-random getallen tussen  $-0.1$  en  $0.1$  bij opgeteld. Ook hier zijn twee gesmoothede reeksen getekend. De reeks met de S-vormige overbrugging van de sprongen heeft overal de gewichten gelijk aan 1. Voor de andere reeks is  $v_i$  nul gesteld op de twee punten vóór en na een sprong, dat wil zeggen  $v_i = v_{i+1} = 0$  als de sprong tussen  $y_i$  en  $y_{i+1}$  optreedt.



Figuur 3.4: Gesimuleerde reeks met sprongen en gesmoothede reeksen met en zonder aangepaste gewichten.

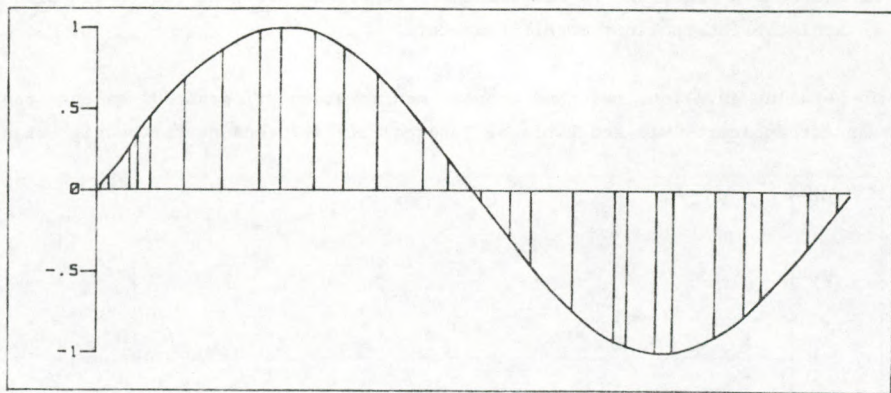


#### 4. Toepassingen voor interpolatie

In het voorafgaande is gebleken dat ontbrekende waarden ingevuld werden door het smoothing-algoritme. Zonder extra inspanningen vond zo interpolatie plaats. We kunnen een stap verder gaan en door een geschikte keus van de reeks  $y_i$  en de gewichten  $w_i$ , het algoritme toespitsen op interpolatie.

In een veel voorkomende situatie zijn equidistant bemonsterde gegevens bekend en wil men geïnterpoleerde waarden berekenen op een  $n$  maal fijner rooster. Het smoothing-algoritme laat zich daar op toepassen door de gegeven waarden in  $y_1, y_{n+1}, y_{2n+1}$  etc. te plaatsen en  $w_1, w_{n+1}, w_{2n+1}$ , etc. 1 te maken; alle andere  $y_i$  en  $w_i$  stelt men nul; alle  $v_i$  stelt men 1. Voor  $\beta$  kiest men een lage waarde, zodat de reeks  $z$  nauwelijks van  $y$  verschilt op de gegeven punten.

Het is niet nodig dat de gegevens op gelijke afstanden over een grof rooster zijn verspreid. Willekeurige afstanden op een fijn rooster zijn ook mogelijk. Het is alleen zaak dat men  $y_i$  invult en de bijbehorende  $w_i$  op 1 stelt. Een voorbeeld geeft figuur 4.1. Daar zijn 27 waarden van  $y_i$  gegeven (de vertikale strepen), op ongelijke afstanden, op een rooster met 180 plaatsen.



Figuur 4.1: Voorbeeld van interpolatie.

Enkele eigenschappen van de interpolerende reeks laten zich, als  $v_i = 1$ , eenvoudig vast stellen aan de hand van de vergelijkingen (2.7). Tussen gegeven punten geldt namelijk  $w_i = 0$ , dus (2.7) vereenvoudigt dan tot:

$$z_{i-2} - 4z_{i-1} + 6z_i - 4z_{i+1} + z_{i+2} = 0 \quad (4.1)$$

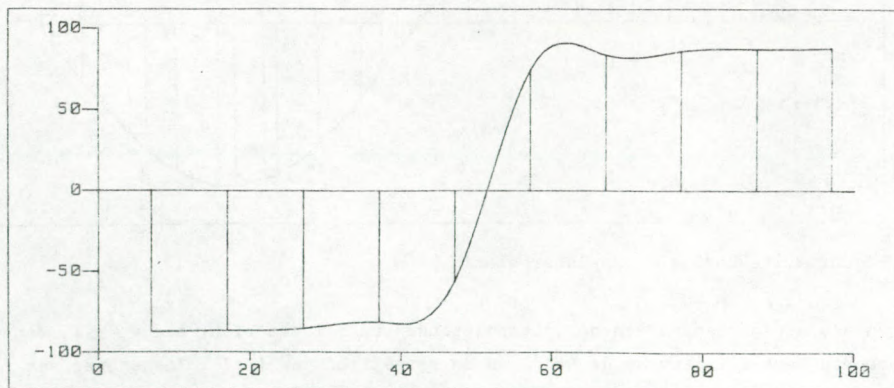
Men kan eenvoudig controleren dat iedere reeks

$$z_i = p_3 i^3 + p_2 i^2 + p_1 i + p_0 \quad (4.2)$$

aan (4.1) voldoet. De interpolerende reeks is dus een derdegraads polynoom. Voor ieder interval tussen gegeven waarden zijn de waarden van  $p_0$  tot en met  $p_3$  verschillend. Een interpretatie van het bovenstaande is dat interpolatie met een spline wordt uitgevoerd. Een interpolerende spline is ook een stuksgewijs derdegraads polynoom (Ahlberg, Nilson en Walsh, 1967). Gebruikelijk is, dat men eerst de coëfficiënten van de polynomen berekent en daaruit de geïnterpoleerde waarden. Het smoothing algoritme doet dat in één keer.

Op analoge wijze kan men aantonen dat toepassen van de eerste differentie tot lineaire interpolatie leidt. Ten opzichte van meer specifieke algoritmen is discrete smoothing voor dit doel omslachtig. Met enkele kleine aanpassingen is echter een algoritme te maken dat een vloeiende overgang tussen lineaire en kubische interpolatie mogelijk maakt.

In bepaalde gevallen, met name wanneer men monotoon stijgende of dalende reeksen interpoleert, wil een kubische interpolatie wel eens te "uitbundig" zijn.



Figuur 4.2: Kubische interpolatie van een monotone reeks.



Er ontstaan dan slingeren die lokaal een niet-monotoon verloop geven. Een voorbeeld is in figuur 4.2 getekend. Voor  $i = 7$ ,  $17.97$  is  $\arctan((i-50)/2)$  gegeven. Vervolgens is geïnterpoleerd met discrete smoothing. Men ziet op beide flanken ongewenste slingeren.

Stel nu dat we de doelfunctie uitbreiden tot:

$$S = \sum_{i=1}^m w_i (y_i - z_i)^2 + \beta v_i \left( \sum_{i=2}^{m-1} (\delta^2 z_i)^2 + \sum_{i=2}^m \gamma (\Delta z_i)^2 \right) \quad (4.3)$$

Uit de partiële afgeleiden volgt:

$$a_i' z_{i-2} - b_i' z_{i-1} + (c_i + w_i) z_i - b_i z_{i+1} + a_i z_{i+2} = w_i y_i \quad (4.4)$$

met:

$$a_i' = \beta v_{i-1} ; \quad b_i' = \beta (2 (v_{i-1} + v_i) + \gamma v_{i-1}) \quad (4.5)$$

$$c_i = \beta ((v_{i-1} + 4 v_i + v_{i+1}) + \gamma (v_{i-1} + v_i))$$

$$a_i = \beta v_{i+1} ; \quad b_i = \beta (2 (v_i + v_{i+1}) + \gamma v_i)$$

Van belang is nu hoe  $z_i$  verloopt tussen gegeven waarden van  $y_i$ , dus als  $w_i = 0$ . Met  $v_i = 1$  voor  $i = 2..m-1$ , levert dat op:

$$z_{i-2} - (4 + \gamma) z_{i-1} + (6 + 2\gamma) z_i - (4 + \gamma) z_{i+1} + z_{i+2} = 0 \quad (4.6)$$

Merk op dat voor  $\gamma = 0$  vergelijking (4.1) ontstaat. Vergelijking (4.6) is een homogene differentievergelijking met constante coëfficiënten. De standaardwijze van oplossen is: stel  $z_i = px^i$ , zie bijvoorbeeld Miller (1966). Na invullen in (4.6) en delen door  $px^i$  ontstaat.

$$x^{-2} - (4 + \gamma) x^{-1} + (6 + 2\gamma) - (4 + \gamma) x + x^2 = 0 \quad (4.7)$$

De coëfficiënten van  $x^{-k}$  zijn hier gelijk aan die van  $x^k$ , dus als  $x_1$  een wortel van (4.7) is, is  $1/x_1$  dat ook. Uit (4.7) volgt:

$$(x - 2 + x^{-1}) (x - 2 - \gamma + x^{-1}) = 0 \quad (4.8)$$

De eerste factor heeft twee identieke wortels die gelijk aan 1 zijn. De grootste wortel van de tweede factor is  $x_2$  (de kleinste is  $1/x_2$ ):

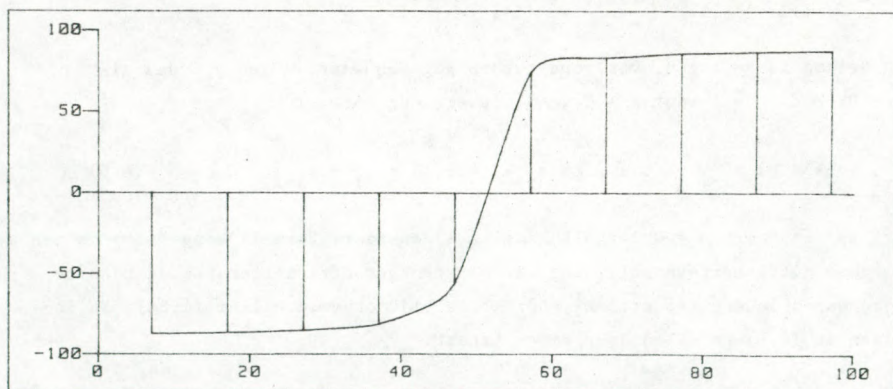
$$x_2 = \frac{2 + \gamma + \sqrt{\gamma(\gamma + 4)}}{2} \quad (4.9)$$

De dubbele wortel 1 draagt aan de oplossing  $p_1 i + p_0$  bij. De twee andere wortels dragen termen  $p_2 x_2^i$  en  $p_3 x_2^{-i}$  bij. De volledige oplossing is dus:

$$z_1 = p_0 + p_1 i + p_2 x_2^i + p_3 x_2^{-i} \quad (4.10)$$

Deze bestaat blijkbaar uit een rechte lijn plus een toenemende en een afnemende exponentiële reeks.

Als illustratie is in figuur 4.3 dezelfde reeks als in figuur 4.2 exponentiëel geïnterpoleerd ( $\gamma = 0.5$ ). De slingeren zijn verdwenen, ten koste van meer geprononceerde knikken. Als curiositeit kan men opmerken dat ook negatieve waarden van  $\gamma$  mogelijk zijn. Dan worden de slingeren in de interpolerende reeks juist versterkt.



Figuur 4.3: Exponentiële interpolatie ( $\gamma = 0.5$ ) van een monotone reeks.



## 5. Discussie

Zoals uit het bovenstaande is gebleken, kan met één, betrekkelijk eenvoudig, algoritme een grote variëteit aan smoothing- en interpolatieproblemen en trendberekeningen uitgevoerd worden. Speciale situaties zoals knikken en sprongen vragen alleen aanpassingen van invoerdata; in het programma zelf zijn daar geen logische beslissingen voor nodig. Ook kan men interpoleren en smoothen met exponentiële functies zonder één  $e$ -macht uit te rekenen. Een nadeel is dat vrij kwistig met geheugenruimte wordt omgesprongen. Zo is het geheugenbeslag het zesvoudige van dat van de ingevoerde te smoothen reeks.

Voor veel smoothing- en interpolatieproblemen maakt men gebruik van splines. Wegman en Wright (1983) geven daar een uitgebreid overzicht van. In de meeste gevallen zal men even goed van discrete smoothing gebruik kunnen maken. De formulering van algoritmen wordt daardoor eenvoudiger.

Toepassingen van discrete smoothing voor het modelleren en smoothen van tijdreeksen vindt men op veel plaatsen in de literatuur, zie bijvoorbeeld Huber (1979), Schlicht (1981) en Akaike (1981). De toepassing van gewichten die nul zijn bij ontbrekende waarnemingen is zeldzaam; voor zover mij bekend hebben alleen Akaike en Ishiguro (1980) er gebruik van gemaakt. Het wegen van de tweede differenties, de expliciete toepassing voor interpolatie en de gewogen combinatie van eerste en tweede differenties in de doelfunctie zijn bij mijn weten nieuw.

## 6. Programmatuur

De hier gegeven voorbeelden zijn allen berekend met een programma van enkele tientallen regels, geschreven in een gestructureerd dialect van BASIC. De geïnteresseerde lezer wende zich tot de auteur.

7. Referenties

- Ahlberg J.H., Nilson E.N. en Walsh J.L. (1967) The Theory of Splines and Their Applications, Academic Press.
- Akaike H. (1981) Seasonal Adjustment by a Bayesian Modelling, Journal of Time Series Analysis 1, 1-13.
- Akaike H. en Ishiguro M. (1980) Trend Estimation with Missing Observations, Ann. Inst. Statist. Math. 32 part B, 481-488.
- Bierman G.J. (1977) Factorization Methods for Discrete Sequential Estimation, Academic Press.
- Cleveland W.S. (1983) Seasonal and Calendar Adjustment, in: Handbook of Statistics, vol. 3, Brillinger D.R. en Krishnaiah P.R. (eds.), North-Holland.
- Fröberg C.E. (1981) Introduction to Numerical Analysis, 2nd ed., Addison-Wesley.
- Henderson R. (1924) A New Method of Graduation, Transactions of the Actuarial Society of America 25, 29-40.
- Huber P.J. (1979) Robust Smoothing, in: Robustness in Statistics, Launer R.L. en Wilkinson G.N. (eds.), Academic Press.
- Miller K.S. (1966) An Introduction to the Calculus of Finite Differences and Difference Equations, Dover.
- Rentrop P. (1980) An Algorithm for the Computation of the Exponential Spline, Numerische Mathematik 35, 81-93.
- Reinsch C.H. (1967) Smoothing by Spline Functions, Numerische Mathematik 10, 177-183.
- Reinsch C.H. (1971) Smoothing by Spline Functions II, Numerische Mathematik 16, 451-454.
- Schlicht E. (1981) A Seasonal Adjustment Principle and Seasonal Adjustment Method Derived from this Principle, J. Am. Stat. Ass. 87, 373-378.
- Wegman E.J., Wright I.W. (1983) Splines in Statistics, J. Am. Stat. Ass. 78, 351-365.
- Whittaker E.T. (1923) On a New Method of Graduation, Proceedings of the Edinburgh Mathematical Society 41, 63-75

Ontvangen: 20-10-1986  
 Geaccepteerd: 13-02-1987