Mathematical study of h-index sequences

Peer-reviewed author version

EGGHE, Leo (2009) Mathematical study of h-index sequences. In: INFORMATION PROCESSING & MANAGEMENT, 45(2). p. 288-297.

DOI: 10.1016/j.ipm.2008.12.002
Handle: http://hdl.handle.net/1942/9279

Mathematical study of h-index sequences

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium[1]

and

Universiteit Antwerpen (UA), IBW, Stadscampus, Venusstraat 35, B-2000 Antwerpen, Belgium

leo.egghe@uhasselt.be

ABSTRACT

This paper studies mathematical properties of h-index sequences as developed by Liang Liming [h-index sequence and h-index matrix: constructions and applications. Scientometrics 69(1), 153-159, 2006]. For practical reasons, Liang Liming studies such sequences where the time goes backwards while it is more logical to use the time going forward (real career periods). Both type of h-index sequences are studied here and their interrelations are revealed. We show cases where these sequences are convex, linear and concave. We also show that, when one of the sequences is convex then the other one is concave, showing that the reverse-time sequence, in general, cannot be used to derive similar properties of the (difficult to obtain) forward time sequence. We show that both sequences are the same if and only if the author produces the same number of papers per year. If the author produces an increasing

---

[1] Permanent address

Key words and phrases: h-index sequence, g-index sequence, R-index sequence, reverse time, Web of Science (WoS).

Acknowledgement: The author thanks M. Goovaerts for his preparation of the Excell sheets used for the raw citation data of WoS and A. Baeten for the preparation of Figs. 1-6.

number of papers per year, then Liang's h-sequences are above the "normal" ones. All these results are also valid for g- and R-sequences. The results are confirmed by the h-, g- and R-sequences (forward and reverse time) of the author.

I. Introduction

In 2005, Hirsch defined his, now famous, Hirsch index or h-index. It was defined in Hirsch (2005) as follows (using our own formulation): if we order the papers of an author in decreasing order of the number of citations received, then the h-index of this author equals h if $r = h$ is the highest rank such that the first h papers each have h or more citations. Since this definition, there has been an "explosion" of papers on the h-index, applying it not only to authors but also to journals (Braun, Glänzel and Schubert (2005, 2006)), to research groups (van Raan (2006)) and even to topics (Banks (2006), The STIMULATE6 Group (2007), Egghe and Rao (2008)).

Advantages and disadvantages of the h-index have been described in the literature (Glänzel (2006a,b), Egghe (2006), Jin, Liang, Rousseau and Egghe (2007)) leading to other indices which have better properties (or at least lack some undesirable properties of the h-index). One obvious disadvantage of the h-index (but also of all other indices) is that it is a fixed number, giving a moment's value of a researcher's career at a certain time. A consequence is also that h-indices of different researchers are difficult to compare (even in the same field) if their career lengths are not the same. Solutions for the latter problem are given in Burrell (2007b) and Jin, Liang, Rousseau and Egghe (2007) but the problem that only one number "describes" a career remains.

We can refer to Egghe (2007a,b,c) and Burrell (2007a) for the first theoretical models for time-dependent h-indices, suggesting concavely increasing h-indices in the papers of Egghe and (approximate) linearly increasing h-indices in function of time, i.e. in function of career length in the Burrell paper.

The problem remains to construct, from year to year, practical h-index sequences of researchers (in short: h-sequences).

This was defined and studied in Liang (2006) but, in our (and Burrell's – see Burrell (2007)) opinion, Liang does not use the most logical definition of a h-sequence. In our opinion, the most logical definition of a h-sequence of a researcher is as follows.

Let the career period of a researcher be described by time $t = 1,2,...,t_m$: here $t = 1$ denotes the first year of the career (more exactly, the year of the first publication) and so on, until $t = t_m$, the final year of the career or the last year we want to cover or the present year (in most cases). Then the h-sequence is constructed as follows. If we only consider the papers of publication year $t = 1$ and their citations obtained in the same year, we then can derive the first h-index, denoted h_1. Next we consider the years $t = 1$ and $t = 2$ together and their citations obtained in the same period, yielding the next h-index, denoted h_2. We continue this way until we reach the final year t_m: we consider all years $t = 1,...,t_m$ and take into account all publications and citations to these publications in this period. This yields the last h-index h_{t_m}. The sequence h_1, h_2, …, h_{t_m} gives a dynamic description of the visibility of this researcher's career and can be compared within the same field, with another researcher's h-sequence.

However, in Liang (2006), another h-sequence is defined: there one uses time in the reverse way (in the direction of the past). Concretely, the first index (which we will denote by h_1^*) is calculated based on the papers published in the year t_m and citations to these papers in the year t_m. The next h-index, denoted h_2^* is calculated, based on the papers published in the years t_m and $t_m - 1$ and the citations to these papers in the same period. We continue this way until we reach year 1: <u>only</u> this h-index (considering all years $t = 1,...,t_m$ for publications as well as citations to these publications), denoted $h_{t_m}^*$ is the same as h_{t_m}.

We underline that the sequence h_1, h_2, …, h_{t_m} is the "natural" h-sequence of a researcher; the sequence h_1^*, h_2^*, …, $h_{t_m}^*$ (denoted without stars in Liang (2006)) was used only for practical reasons: (only) if t_m is the present year, one can calculate h_1^*, …, $h_{t_m}^*$ in an automatic way from the Web of Science (WoS). In the WoS, only citation data, and subsequent h-indices are given (whatever the set of articles) for the citing period up to the present year. That is why Liang calculated h_1^*, …, $h_{t_m}^*$ instead of the more natural h_1, …, h_{t_m} for which one has to collect all citation data from the WoS and to restrict the citing period ($t = 1$, $t = 1$ and $t = 2$, …) manually, which is very time-consuming.

We fully understand that Liang wanted to avoid the time-consuming calculation of the sequence $h_1, \ldots, h_{t_m}$ (for eleven physicists) by replacing it by the sequence $h_1^*, \ldots, h_{t_m}^*$ but, in this case, we need to know that the latter sequence resembles the former one. The comparison of both sequences, in a logical Lotkaian publication-citation environment, is the topic of this paper.

The h-sequences will be studied for continuous time $t \in \mathbb{R}^+$. Also we will suppose that, for each time period (backwards or forward), we have an information production process (IPP) (cf. Egghe (2005)) of publications and citations to these publications conforming with Lotka's law

$$f(j) = \frac{C}{j^\alpha} \qquad (1)$$

$C > 0$, $\alpha > 1$, where $f(j)$ denotes the density of the articles with a density j of citations to these articles (see Egghe (2005)). We assume that α is constant in each time period considered. It is clear that this simplification does not jeopardises the conclusions of this paper concerning the comparison of both h-sequences. Since we take time as a continuous variable we will denote the sequence $h_1, \ldots, h_{t_m}$ by $h(t)$, $t \in [0, t_m]$ and the sequence $h_1^*, \ldots,$ $h_{t_m}^*$ by $h^*(t)$, $t \in [0, t_m]$.

In the next section we will study $h(t)$ for a fixed number of publications per time unit (say per year) and for an increasing number of publications per time unit, where the increase is expressed using a power function or an exponential function. Necessary and sufficient conditions are given for the h-"sequence" (function) $h(t)$ to be convexly, linearly or concavely increasing and we indicate that the concave increase is the most natural one.

In the third section we define the h^*-function $h^*(t)$ in function of $h(t)$ and prove a necessary and sufficient condition for $h^*(t) = h(t)$ for all $t \in [0, t_m]$, the ideal situation: indeed, only in this case we can substitute $h^*(t)$ (the one that can be calculated in an automatic way) for $h(t)$ (the one that requires a lot of manual intervention but the natural one). It turns out that $h(t) = h^*(t)$, for all $t \in [0, t_m]$ and all $t_m > 0$, if and only if the number of publications per time unit (say a year) of the researcher is constant. This is an important case but, as shown by the author's data, is only a rough approximation of reality. We also show that $h^*(t) \geq h(t)$, for all $t \in [0, t_m]$ if the number of publications per time unit (year) of the researcher increases.

In the fourth section we even prove that, for general publication production schemes, if one of the functions $h(t)$ or $h^*(t)$ is convex (including the linear case), then the other one is concave, showing that, in general, the behavior of the h-"sequences" $h(t)$ and $h^*(t)$ is different. We also note that the converse of the above assertion is false by giving examples of cases where both $h(t)$ and $h^*(t)$ are concave.

The fifth section gives $h(t)$ and $h^*(t)$ for this author. It is also remarked that the same results hold for the g-index (Egghe (2006)) and the R-index (Jin, Liang, Rousseau and Egghe (2007)) and are illustrated by presenting $g(t)$, $g^*(t)$, $R(t)$ and $R^*(t)$ for this author.

We can then conclude that one cannot avoid the time-consuming task of calculating the natural h-sequence $h_1, \ldots, h_{t_m}$ (and similarly for the g- and R-index) of a researcher except in the case the researcher has a (more or less) constant publication production per year. We end the paper by proposing open problems and advises.

II. Study of the h-"sequence" h(t)

For the natural h-sequence (function) $h(t)$, we consider the career of a researcher from the start $(t = 0)$ up to a time $t > 0$. Let us denote by $T(t)$ the total number of publications of this researcher at time t. This set of publications is assumed to have citations (in the same period $[0, t]$) according to Lotka's law (1), with α independent from t (as explained in the

Introduction). It was proved in Egghe and Rousseau (2006) that, in this case, the h-index (t-dependent here) equals

$$h(t) = T(t)^{\frac{1}{\alpha}} \quad (2)$$

for each $t \in [0, t_m]$ (note that, for $t = 0$, we have $T(0) = 0$ and $h(0) = 0$ naturally).

We will now study the shape of the function $h(t)$ in three simple, natural cases.

II.1 The case of constant production

If a researcher publishes the same number of papers per time unit (e.g. a year), say b, then we have that, for every $t \in [0, t_m]$ that $T(t) = bt$. Then (2) implies

$$h(t) = b^{\frac{1}{\alpha}} t^{\frac{1}{\alpha}} \quad (3)$$

which is a concavely increasing function since $\alpha > 1$. This is the simplest model for $h(t)$ and is a first approximation of reality.

Table 1 shows this author's yearly production of publications (articles and books) according to publication year. Time $t = 1$ is the starting year 1978, up to 2007, totalling 30 years of publications.

Table 1. Number of publications per year of L. Egghe

t	#	t	#	t	#	t	#	t	#
1	2	7	3	13	8	19	8	25	8
2	0	8	4	14	5	20	11	26	9
3	4	9	3	15	8	21	8	27	9
4	5	10	5	16	4	22	10	28	13
5	5	11	9	17	5	23	11	29	15
6	2	12	0	18	6	24	7	30	14

It is clear that a constant production per year is not the case. One can see a moderate increase which can be described by a power function or an exponential function. These cases will be studied below.

II.2 The case of increasing production per year, using a power function

Here we assume a number (density) of publications per time unit being bt^{β} where b, $\beta > 0$ (the case $\beta = 0$ corresponds to the previous case). Hence for every $t \in [0, t_m]$:

$$T(t) = \int_0^t bt'^{\beta} dt' = \frac{b}{\beta + 1} t^{\beta + 1} \tag{4}$$

Now, according to (2) we have

$$h(t) = \left(\frac{b}{\beta + 1} \right)^{\frac{1}{\alpha}} t^{\frac{\beta + 1}{\alpha}} \tag{5}$$

which is concave iff $\beta + 1 < \alpha$, linear iff $\beta + 1 = \alpha$ and convex iff $\beta + 1 > \alpha$. One can expect that a researcher's production does not increase fastly so that a small β occurs more often in which case we again expect a concavely increasing function $h(t)$.

II.3 The case of increasing production per year, using an exponential function

Here we assume a number (density) of publications per time unit being bc^t where $b > 0$, $c > 1$. One could even take $0 < c < 1$ in which we have a decreasing yearly production. The case $c = 1$ corresponds to the case studied in subsection II.1. Now, for every $t \in [0, t_m]$:

$$T(t) = \int_0^t bc^{t'} dt' = \frac{b}{\ln c} \left(c^t - 1 \right) \tag{6}$$

Now, according to (2) we have

$$h(t) = \left(\frac{\beta}{\ln c}\right)^{\frac{1}{\alpha}} \left(c^t - 1\right)^{\frac{1}{\alpha}} \tag{7}$$

which is, for $c > 1$, increasing and where $h''(t) < 0$ iff $c^t < \alpha$. So here, dependent on the values of c, α and t we can have a concave $h(t)$ or an S-shaped $h(t)$ (since $\alpha > 1$ we have that, for t small enough we always have $c^t < \alpha$ so that a completely convex $h(t)$ is not possible here).

For a moderate increase per year of the production (i.e. $c > 1$ but close to 1) we hence have, if t_m is not very large, that $h(t)$ is concavely increasing.

We now start the study of the reverse function $h^*(t)$. The next section defines this function and presents a necessary and sufficient condition for $h^*(t) = h(t)$ for all $t \in [0, t_m]$.

III. Liang's h-sequence $h^*(t)$

Let us denote the career length of a researcher by t_m: hence the researcher has publications in the time period $[0, t_m]$. Let us denote, as in the previous section, by $T(t)$ the total number of publications of this researcher at time t (i.e. t time units since the start of the career at $t = 0$). Liang starts at time t_m, going back to the past as depicted in Fig. 1: For Liang, time t is the period (in normal time) between $t_m - t$ and t_m (for $t \le t_m$).

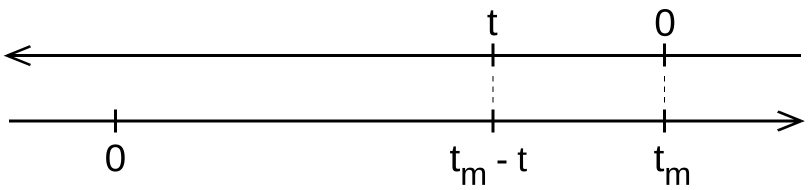

Fig. 1. Time and reverse time

Hence, in reverse time, one considers a number of publications, denoted as $T^*(t)$, equalling

$$T^*(t) = T(t_m) - T(t_m - t) \tag{8}$$

In the Introduction we assumed that, for any time t in the direct way or in the reverse way, the number of publications $T(t)$ or $T^*(t)$ have citations according to Lotka's law (1), where $\alpha > 1$ is fixed. This yields the following Proposition.

Proposition III.1:

Let $h(t)$ denote the h-sequence of a researcher for normal time and $h^*(t)$ denote the h-sequence of this researcher for the reverse time as described above (e.g. formula (8)). Then we have, supposing Lotka's law (1),

$$h^*(t) = \left(T(t_m) - h(t_m - t)^\alpha\right)^{\frac{1}{\alpha}} \tag{9}$$

for all $t \in [0, t_m]$.

Proof:

Using Lotka's law (1) we have that, for every $t \in [0, t_m]$

$$h(t) = T(t)^{\frac{1}{\alpha}}$$

, by (2). If we apply this for $t_m - t$ (also belonging to the interval $[0, t_m]$) we have

$$h(t_m - t) = T(t_m - t)^{\frac{1}{\alpha}}$$

or

$$T(t_m - t) = h(t_m - t)^\alpha \tag{10}$$

Further, using again (1) and (2) to the publication set $T^*(t)$, we have that

$$h^*(t) = T^*(t)^{\frac{1}{\alpha}} \tag{11}$$

for all $t \in [0, t_m]$. Formulae (8), (10) and (11) prove formula (9), finishing this proof. □

Although the Liang h-“sequence” $h^*(t)$ has interest in itself, it can only give information about the natural h-sequence $h(t)$ if they are (more or less) equal. A characterization of this will be given in the next Theorem.

Theorem III.2:
Both h-sequences $h(t)$ and $h^*(t)$ are identical:

$$h(t) = h^*(t) \tag{12}$$

for all $t \in [0, t_m]$ and all $t_m \in \mathbb{R}^+$ if and only if the researcher has a constant production of publications per time unit. In other words: (12) is valid iff

$$T(t) = bt \tag{13}$$

for a certain constant $b > 0$.

Proof: Formula (9) and (12) yield, for all $t \in [0, t_m]$

$$h^*(t) = \left(T(t_m) - h(t_m - t)^\alpha\right)^{\frac{1}{\alpha}}$$

$$= h(t) = T(t)^{\frac{1}{\alpha}},$$

using also (2). Using again (2) we have

$$h(t_m - t)^\alpha = T(t_m - t)$$

so that we have (necessary and sufficient to have (12))

$$T(t_m) - T(t_m - t) = T(t) \tag{14}$$

(otherwise stated: $T^*(t) = T(t)$ for all $t \in [0, t_m]$, by (8)).

Denoting $t = x$, $t_m - t = y$, hence $t_m = x + y$, (14) requires

$$T(x + y) = T(x) + T(y) \tag{15}$$

for all $x, y \in \mathbb{R}^+$ (since the above is required for all $t \in [0, t_m]$ and all $t_m > 0$). Relation (15) can be extended to all $x, y \in \mathbb{R}$ by defining, for $x \in \mathbb{R}^-$: $T(x) = T(-x)$ so that (15) is valid for all $x, y \in \mathbb{R}$. Since we, evidently, assume that the function $T(.)$ is continuous, we have, by a well-known result (cf. Roberts (1979) – see also Egghe (2005), Appendix 1, Theorem A.I.1) that the function $T(.)$ must be linear: there exists a number $b \in \mathbb{R}$ such that

$$T(t) = bt \tag{16}$$

Of course, since $T(t) > 0$ for all $t > 0$, we have $b > 0$, completing the proof of this Proposition. □

Note that (13) trivially implies (12) since $h(t) = T(t)^{\frac{1}{\alpha}} = (bt)^{\frac{1}{\alpha}}$ and since, by (8):

$T^*(t) = T(t_m) - T(t_m - t) = bt_m - b(t_m - t) = bt$, hence $h^*(t) = T^*(t)^{\frac{1}{\alpha}} = (bt)^{\frac{1}{\alpha}} = h(t)$, for all $t \in [0, t_m]$.

The case that a researcher has a constant number of publications per time unit is an important simple case and a first approximation of reality: we can indeed, roughly, assume that a researcher, in his/her career, produces more or less the same number of papers per year, certainly in the middle part of the career: in the beginning of the career, the researcher will produce (most probably) less papers (cf. Table 1) as is probably also the case at the end of a career (luckily I cannot illustrate this yet !).

But the above Proposition also indicates that, in all cases where paper production is not constant per time unit, we have that $h^*(t) \neq h(t)$.

Could it then be that the shapes of both functions are the same ? Then the (easier) calculation of $h^*(t)$ could at least give some insight in the <u>shape</u> of $h(t)$, which is, as said above, very time-consuming to calculate. The next section shows that, in general, the shapes of $h^*(t)$ and $h(t)$ are very different, jeopardizing, to a large extent, the use of $h^*(t)$ as a substitute for $h(t)$.

We close this section with the following Theorem on the comparison of $h^*(t)$ and $h(t)$.

<u>Theorem III.3</u>:
If $T'(t)$ (strictly) increases (i.e. $T(t)$ convex) then

$$h^*(t) \geq (>) h(t)$$

for every $t \in [0, t_m]$.

<u>Proof</u>:
It follows from (2) and (9) that $h^*(t) > h(t)$ if and only if

$$T(t_m) > T(t_m - t) + T(t) \tag{17}$$

(and similarly for the $\geq$ sign; we leave this to the reader).

<u>(i) Let $t \geq \frac{t_m}{2}$</u>

By the mean value theorem on the function $T(t)$ (supposed to be differentiable) we have

$$T(t_m) - T(t) = T'(\lambda)(t_m - t) \tag{18}$$

for a certain $\lambda \in \left]t, t_m\right[$ and

$$T(t_m - t) = T(t_m - t) - T(0)$$

$$= T'(\xi)(t_m - t) \tag{19}$$

for a certain $\xi \in \left]0, t_m - t\right[$.

Since $t > \frac{t_m}{2}$ we have that $\left]t, t_m\right[\subset \left]\frac{t_m}{2}, t_m\right[$ and $\left]0, t_m - t\right[\subset \left]0, \frac{t_m}{2}\right[$. Hence, since $T'(t)$ strictly increases, we have $T'(\xi) < T'(\lambda)$. Hence (18) and (19) imply

$$T(t_m) - T(t) > T(t_m - t)$$

hence (17).

<u>(ii) Let $t < \frac{t_m}{2}$</u>

By the mean value theorem on $T(t)$ we now have

$$T(t_m) - T(t_m - t) = T'(\eta)t \tag{20}$$

for a certain $\eta \in \left]t_m - t, t_m\right[$ and

$$T(t) = T(t) - T(0)$$

$$= T'(\kappa)t \tag{21}$$

for a certain $\kappa \in \left]0,t\right[$. Now, since $t < \frac{t_m}{2}$ we have that $\left]t_m - t, t_m\right[\subset \left[\frac{t_m}{2}, t_m\right]$ and $\left]0,t\right[\subset \left[0, \frac{t_m}{2}\right]$. Hence, since $T'(t)$ strictly increases, we have by (20) and (21) $T'(\kappa) < T'(\eta)$ and hence

$$T(t_m) - T(t_m - t) > T(t)$$

yielding again (17). This concludes the proof. □

Note that $T'(t)$ denotes the number of papers per time unit, say, in the discrete case, the production per year. So Theorem III.3 deals with an increasing number of papers, say per year. Note that the conclusion of theorem III.3 ($h^*(t) > h(t)$ for all t) implies that $h^*(t)$ cannot be convex in this case (since $h^*(0) = h(0) = 0$ and $h^*(t_m) = h(t_m)$).

Of course, if $T'(t)$ (strictly) decreases (what we do not expect to be the case, usually) we have $h^*(t) \leq (<) h(t)$ for every $t \in [0, t_m]$. In this case, $h(t)$ cannot be convex.

Note that $h^*(0) = h(0) = 0$ and $h^*(t_m) = h(t_m) = T(t_m)^{\frac{1}{\alpha}}$ (by (2) and (9)). The tangent lines in $t = 0$ have slopes $h^{*\prime}(0)$ and $h'(0)$. We have

$$\lim_{\substack{t \to 0 \\ >}} \frac{h^{*\prime}(t)}{h'(t)} = \lim_{\substack{t \to 0 \\ >}} \frac{h^*(t)}{h(t)} \tag{22}$$

by de l'Hôspital's rule (and since $h^*(0) = h(0) = 0$ and since $h^*(t)$ and $h(t)$ are continuous). By (2) and (9) we have

$$\lim_{\substack{t \to 0 \\ >}} \frac{h^*(t)}{h(t)} = \lim_{\substack{t \to 0 \\ >}} \frac{T(t_m) - T(t_m - t)}{T(t)}$$

$$= \frac{T'(t_m)}{T'(0)} \tag{23}$$

again by the l'Hôspital's rule. In the case of Theorem III.3 is the highest difference between the values $T'(t)$, $t \in [0, t_m]$ given by $T'(0)$ (smallest value) and $T'(t_m)$ (largest value), so that (22) and (23) imply that

$$\lim_{\substack{t \to 0 \\ >}} \frac{h^{*}{'}(t)}{h'(t)} > 1$$

which means that, from $t = 0$ onwards, $h^{*}(t)$ increases faster than $h(t)$.

IV. General relations between h(t) and h*(t)

We can prove the following result which is bad news for the useability of the $h^{*}(t)$ function instead of the function $h(t)$.

Theorem IV.1:
In the general situation of Proposition III.1, we have that, if $h(t)$ is convex (including the linear case), then $h^{*}(t)$ is strictly concave.

Proof: Based on (9) we have the general formulae, for all $t \in [0, t_m]$

$$h^{*}{'}(t) = \left(T(t_m) - h(t_m - t)^{\alpha}\right)^{\frac{1}{\alpha} - 1} h(t_m - t)^{\alpha - 1} h'(t_m - t)$$

which is strictly positive (assuming, of course, that $T(t)$ strictly increases) since $h(t_m - t)^{\alpha} = T(t_m - t) < T(t_m)$ and since $h' > 0$ (by (2) and the fact that $T(t)$ strictly increases). Further

$$h^{*}{}''(t)=\left(\frac{1}{\alpha}-1\right)\left(T(t_m)-h(t_m-t)^{\alpha}\right)^{\frac{1}{\alpha}-2}\alpha h(t_m-t)^{2\alpha-2}.h'^{2}(t_m-t)$$

$$+\left(T(t_m)-h(t_m-t)^{\alpha}\right)^{\frac{1}{\alpha}-1}(\alpha-1)h(t_m-t)^{\alpha-2}h'^{2}(t_m-t)(-1)$$

$$+\left(T(t_m)-h(t_m-t)^{\alpha}\right)^{\frac{1}{\alpha}-1}h(t_m-t)^{\alpha-1}h''(t_m-t)(-1).$$

Each term in the above formula is negative because $\alpha>1$, since $h''(t)\geq 0$ for all $t\in[0,t_m]$ and again since $h(t_m-t)^{\alpha}<T(t_m)$. This shows the strict concavity of h^{*} (i.e. $h^{*}{}''(t)<0$ for all $t\in[0,t_m]$) (even in case $h(t)$ is linear). □

If we replace $h(t)$ and $h^{*}(t)$ in the above Theorem, we again have a valid result.

<u>Theorem IV.2</u>:

In the general situation of Proposition III.1, we have that, if $h^{*}(t)$ is convex (including the linear case), then $h(t)$ is strictly concave.

<u>Proof</u>:

We again invoke equation (9):

$$h^{*}(t)=\left(T(t_m)-h(t_m-t)^{\alpha}\right)^{\frac{1}{\alpha}}$$

, valid for all $t\in[0,t_m]$. A little bit of algebra yields, for all $t\in[0,t_m]$

$$h(t_m-t)=\left(T(t_m)-h^{*}(t)^{\alpha}\right)^{\frac{1}{\alpha}}$$

for all $t\in[0,t_m]$. Putting $x=t_m-t$ we have: for all $x\in[0,t_m]$

$$h(x) = \left(T(t_m) - h^*(t_m - t)^\alpha\right)^{\frac{1}{\alpha}}$$

which is exactly formula (9) but for $h(.)$ and $h^*(.)$ reversed. Hence the Theorem follows from Theorem IV.1. □

We next show, by one example, that the converses of Theorems IV.1 and IV.2 are not true. This is done by calculating $h^*(t)$ for the case studied in Subsections II.1 and II.2.

<u>Example : h*(t) versus h(t) in case of Subsection II.1</u>

Here we assumed a constant number of papers per time unit and we obtained (Subsection II.1 and Theorem III.2) for all $t \in [0, t_m]$:

$$h(t) = b^{\frac{1}{\alpha}} t^{\frac{1}{\alpha}} = h^*(t),$$

hence both $h(t)$ and $h^*(t)$ are strictly concave.

<u>Example : h*(t) versus h(t) in case of Subsection II.2</u>

Here we assumed (4) yielding formula (5). From these results we have, using formula (9) that

$$h^*(t) = \left(\frac{b}{\beta + 1}\right)^{\frac{1}{\alpha}} \left(t_m^{\beta+1} - (t_m - t)^{\beta+1}\right)^{\frac{1}{\alpha}} \tag{24}$$

A quick calculation shows that $h^*(t)$ is <u>always</u> strictly concave since $\alpha > 1$. Note that we showed in Subsection II.2 that $h(t)$ is concave iff $\beta + 1 < \alpha$ (yielding the case that both $h(t)$ and $h^*(t)$ functions are concave), that $h(t)$ is linear iff $\beta + 1 = \alpha$ and that $h(t)$ is convex iff $\beta + 1 > \alpha$ (the last two cases confirm Theorem IV.1).

V. Verification of these results on the h-sequences of Egghe and extension to the g- and R-sequences

Fig. 2 gives the h-sequence of this author where $t = 1$ is the year 1978. So $h(1)$ is the h-index for the set publications in 1978 and citations to these papers in 1978. Similarly $h(2)$ is the h-index for the set publications in the 1978-1979 period and citations in this period to these papers, and so on: $h(30)$ is the h-index for the set publications in the 1978-2007 period and citations to these papers in this period.

Fig. 3 gives the h^*-sequence of this author, where now $t = 1$ is 2007. So $h^*(1)$ is the h-index for the set of publications in 2007 and citations to these papers in 2007. Similarly $h^*(2)$ is the h-index for the set publications in the period 2006-2007 and citations in this period to these papers, and so on: $h^*(30)$ is the h-index for the set publications in the 1978-2007 period and citations to these papers in this period. So, only for $t = 30$ we have $h(30) = h^*(30)$.

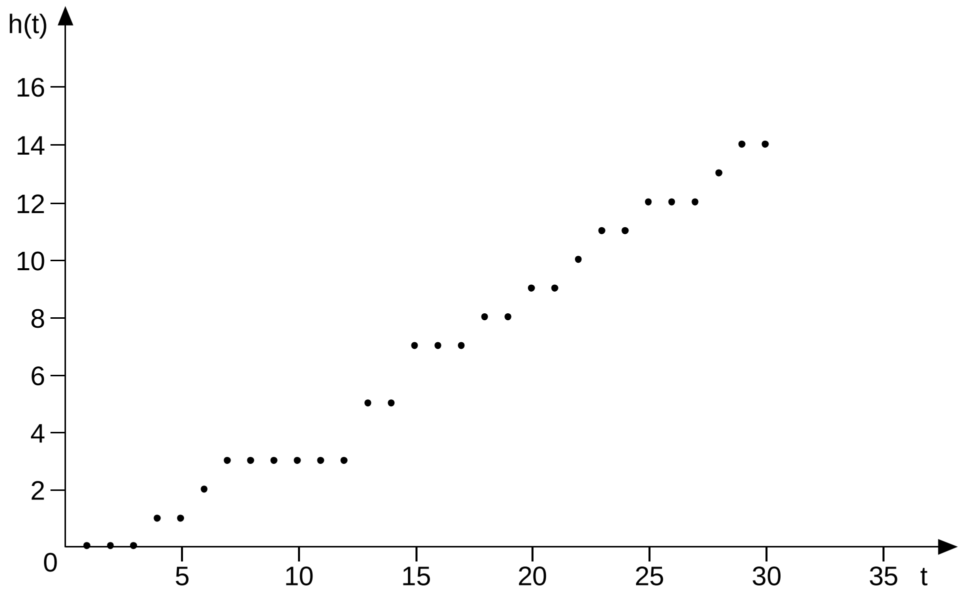

Fig. 2. $h(t)$ for the career of Egghe

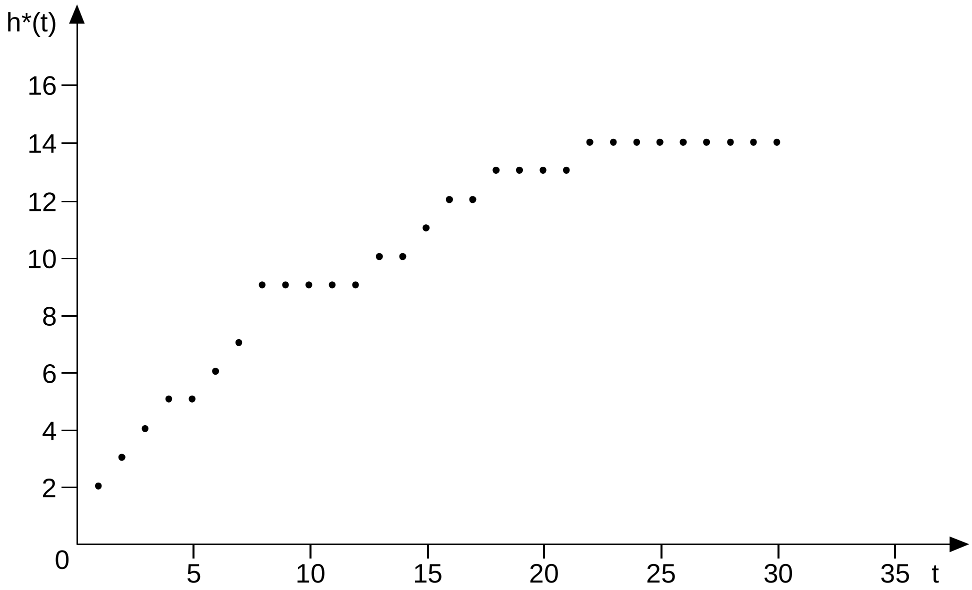

Fig. 3. $h^*(t)$ for the career of Egghe

We clearly see that $h(t)$ is approximately linear (confirming in this case Burrell (2007a)) while $h^*(t)$ has a more concave shape, confirming Theorem IV.1. We also see that $h^*(t) > h(t)$ confirming Theorem III.3 taking into account the “overall” increase of the number of papers per year (Table 1). We also see that, for small t, $h(t)$ increases more slowly than $h^*(t)$, also confirming results of Section III. This is also logical since, in the beginning of a career, the first papers will, normally, be cited in a later period. Conversely, $h^*(t)$ increases more slowly, for larger t, than $h(t)$ which is also logical since high t for $h^*(t)$ means: the beginning period.

As noted in Egghe (2006) and Jin, Liang, Rousseau and Egghe (2007) the indices g and R are fixed (for α fixed) multiples of h. Hence the theoretical results, proved in Section II, III and IV are also valid for the sequences $g(t)$, $R(t)$(forward time) and $g^*(t)$ and $R^*(t)$ (reverse time).

For each period, the g-index was defined as the largest rank such that the total number of citations to the papers on ranks $1,\ldots,g$ is larger than or equal to g^2. Also, R equals the square root of the sum of the number of citations to the first h papers. They were introduced to avoid the disadvantage of the h-index that it is insensitive to the number of citations to papers in the h-core.

That the results on $h(t)$ and $h^*(t)$ are also valid for $g(t)$, $g^*(t)$, $R(t)$ and $R^*(t)$ is illustrated in Figs. 4, 5, 6 and 7 based on the same data of Egghe used in Figs. 2, 3.

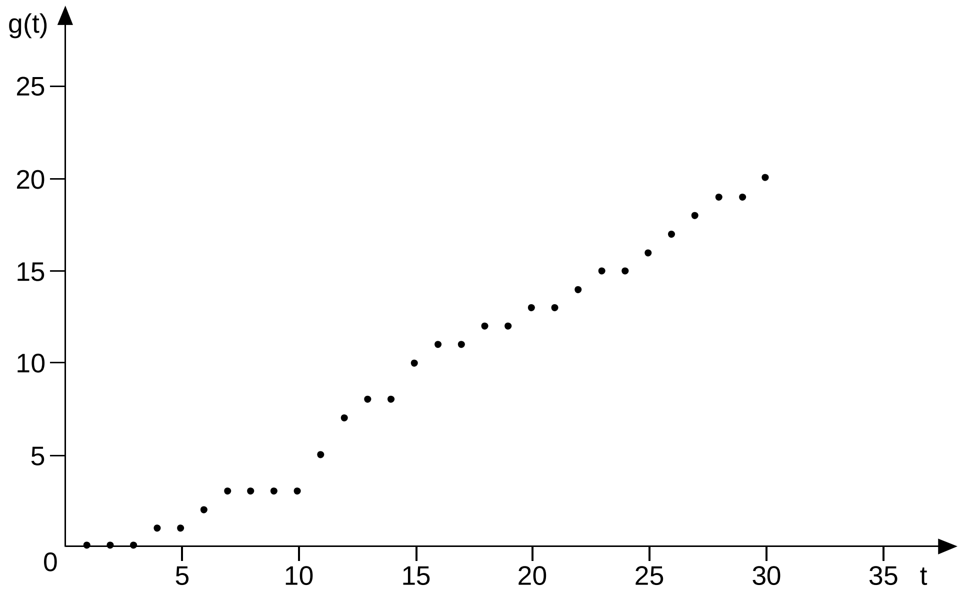

Fig. 4. $g(t)$ for the career of Egghe

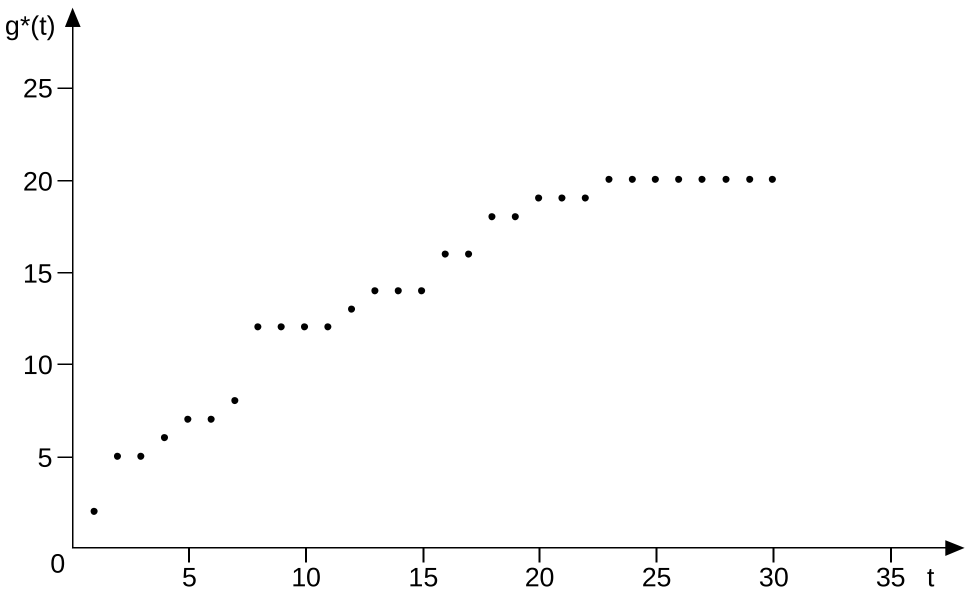

Fig. 5. $g^*(t)$ for the career of Egghe

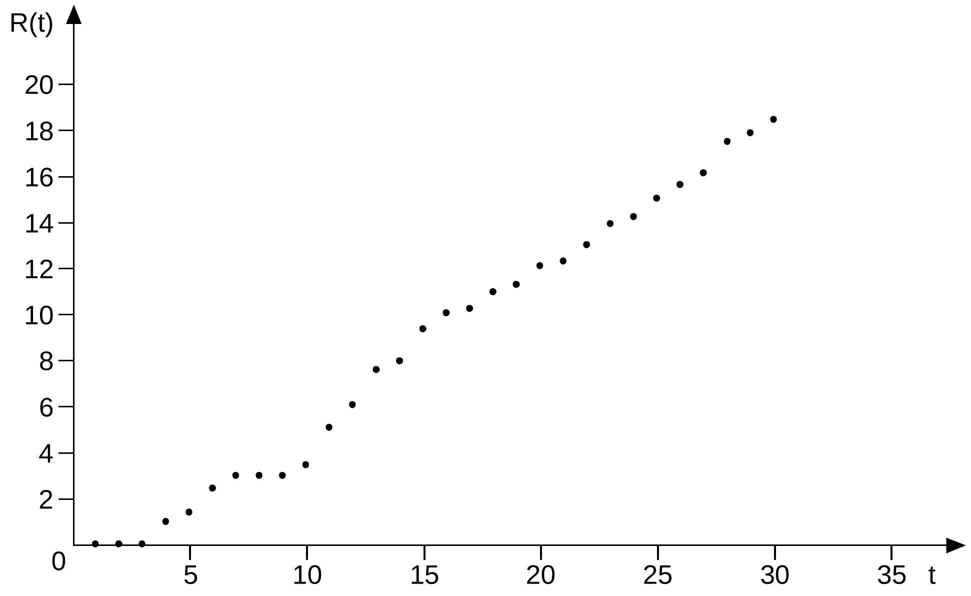

Fig. 6. $R(t)$ for the career of Egghe

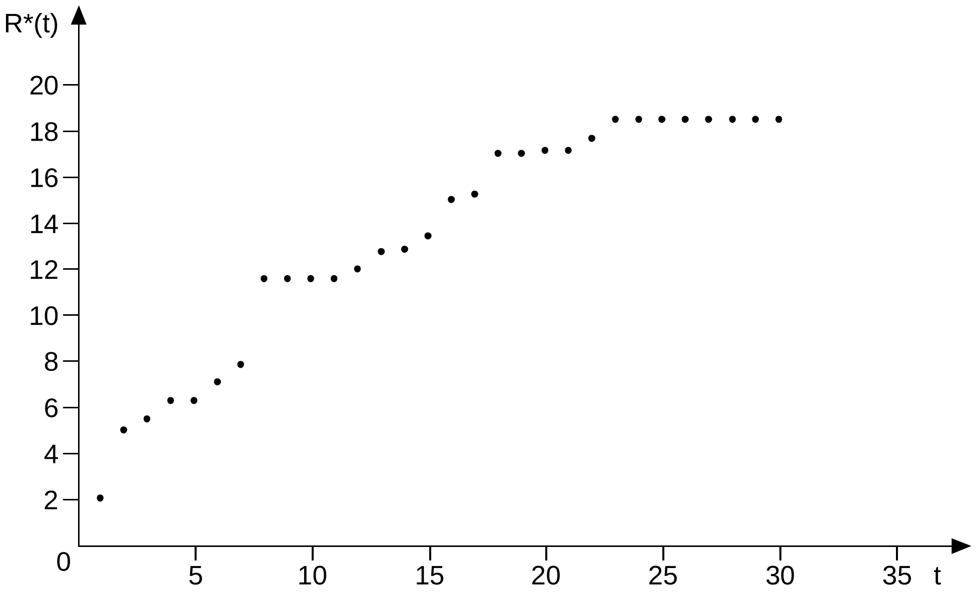

Fig. 7. $R^{*}(t)$ for the career of Egghe

VI. Remarks and conclusions

This paper studied the sequences (functions) $h(t)$ (forward time) and $h^*(t)$ (reverse time, as defined by Liang). Concrete examples are given based on paper production models per time unit. We showed that $h(t)= h^*(t)$ for all t, iff paper production per time unit is constant and that $h^*(t)> h(t)$ for all t if the paper production per time unit strictly increases. We also show that even the shapes of $h(t)$ and $h^*(t)$ can be very different (e.g. one being convex and the other one being concave). Hence, we can conclude that, apart from constant paper production, $h^*(t)$ cannot be used to study $h(t)$ which is a pity since $h^*(t)$ can be generated from the WoS in an automatic way while $h(t)$ (the sequence which is the more natural one) must be calculated in a time-consuming manual way.

It is our advise – as also Burrell did (Burrell (2007b)) – that the WoS produces such h-sequences in an automatic way. It would be very good that researchers and policy makers can have h-sequences at their disposition: it better shows the evolution of a career which a single h-index cannot. Of course, from now on, a researcher can calculate his/her h-index on a yearly basis, generating a h-sequence after many years, but this, obviously, requires a lot of patience!

We remarked that all theoretical result on $h(t)$ and $h^*(t)$ are also true for the g-index and R-index.

We presented $h(t)$, $h^*(t)$, $g(t)$, $g^*(t)$, $R(t)$ and $R^*(t)$ for the career of this author (period 1978-2007) and confirm hereby the theoretical results.

We obtained a linearly increasing $h(t)$ function for the Egghe data. Yet, the formula (2) shows a concave (since $\alpha > 1$) relation of h in function of T. As proved in Section II, in many models is $h(t)$ a concave function of time t. We can conclude in the Egghe case that the cumulative production $T(t)$ more or less compensates the exponent $\frac{1}{\alpha}$ in (2) so that (approximately) $T(t)= ct^{\alpha}$ (c a constant) resulting in

$$h(t) = T(t)^{\alpha} = \left(ct^{\alpha}\right)^{\frac{1}{\alpha}}$$

$$= c^{\frac{1}{\alpha}} t$$

, a linear function of t. In this case, the function $h(t)$ follows the straight line connecting $(0,0)$ and $(t_m, h(t_m))$, which has equation

$$y = \frac{h(t_m)}{t_m} t \qquad (25)$$

The function $h(t)$ is entirely below this line if and only if

$$h(t) = T(t)^{\frac{1}{\alpha}} \leq \frac{h(t_m)}{t_m} t$$

which is equivalent with

$$T(t) \leq T(t_m) \left(\frac{t}{t_m} \right)^{\alpha} \qquad (26)$$

Note that in this case $h(t)$ cannot be concave (as e.g. is the case for $T(t) = bt$ for all t, which does not satisfy (26) since $\alpha > 1$).

Similar inequalities can be given for $h(t)$ above (25) and $h^{*}(t)$ below or above (25).

References

M.G. Banks (2006). An extension of the Hirsch index: indexing scientific topics and compounds. Scientometrics 69(1), 161-168.

T. Braun, W. Glänzel and A. Schubert (2005). A Hirsch-type index for journals. The Scientist, 19(22), 8.

T. Braun, W. Glänzel and A. Schubert (2006). A Hirsch-type index for journals. Scientometrics 69(1), 169-173.

Q.L. Burrell (2007a). Hirsch's h-index: a stochastic model. Journal of Informetrics 1(1), 16-25.

Q.L. Burrell (2007b). Hirsch index or Hirsch rate ? Some thoughts arising from Liang's data. Scientometrics 73(1), 19-28.

L. Egghe (2005). Power Laws in the Information Process: Lotkaian Informetrics. Elsevier, Oxford (UK).

L. Egghe (2006). Theory and practice of the g-index. Scientometrics 69(1), 131-152.

L. Egghe (2007a). Dynamic h-index: the Hirsch index in function of time. Journal of the American Society for Information Science and Technology 58(3), 452-454.

L. Egghe (2007b). Item-time-dependent Lotkaian informetrics and applications to the time-dependent h- and g-index. Mathematical and Computer Modelling 45(7-8), 864-872.

L. Egghe (2007c). Time-dependent Lotkaian informetrics incorporating growth of sources and items. Mathematical and Computer Modelling, to appear.

L. Egghe and I.K.R. Rao (2008). The influence of the broadness of a query of a topic on its h-index: Models and examples of the h-index of N-grams. Preprint.

L. Egghe and R. Rousseau (2006). An informetric model for the Hirsch-index. Scientometrics 69(1), 121-129.

W. Glänzel (2006a). On the opportunities and limitations of the h-index. Science Focus 1(1), 10-11 (in Chinese).

W. Glänzel (2006b). On the h-index – a mathematical approach to a new measure of publication activity and citation impact. Scientometrics 67(2), 315-321.

J.E. Hirsch (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America 102, 16569-16572.

B. Jin, L. Liang, R. Rousseau and L. Egghe (2007). The R- and AR-indices : Complementing the h-index. Chinese Science Bulletin 52(6), 855-863.

L. Liang (2006). h-index sequence and h-index matrix: Constructions and applications. Scientometrics 69(1), 153-159.

F.S. Roberts (1979). Measurement Theory with Applications to Decisionmaking, Utility and the Social Sciences. Addison-Wesley, Reading (MA), USA.

The STIMULATE6 Group (2007). The Hirsch index applied to topics of interest to developing countries. First Monday 12(2), 2007. http://www.firstmonday.org/issues/issues/issues12 2/stimulate/

A.F.J. van Raan (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups. Scientometrics 67(3), 491-502.