A lower bound on the dimension of positive realizations

Béla Nagy, Máté Matolcsi

Béla Nagy is Professor at the Mathematics Department of Technical University, Budapest. E-mail: bnagy@math.bme.hu

Máté Matolcsi is a research assistant in Applied Mathematics at the Alfred Renyi Institute, Budapest. E-mail: matomate@renyi.hu

Abstract

A basic phenomenon in positive system theory is that the dimension N of an arbitrary positive realization of a given transfer function $H(z)$ may be strictly larger than the dimension n of its minimal realizations. The aim of this brief is to provide a non-trivial lower bound on the value of N under the assumption that there exists a time instant k_0 at which the (always nonnegative) impulse response of $H(z)$ is 0 but the impulse response becomes strictly positive for all $k > k_0$. Transfer functions with this property may be regarded as extremal cases in positive system theory.

Keywords

Positive linear systems, positive realization, dimension estimates

I. Introduction

Let the transfer function of a discrete time-invariant linear scalar system

$$H(z) = \frac{p_1 z^{n-1} + ... + p_n}{z^n + q_1 z^{n-1} + ... + q_n} = \sum_{k=1}^{\infty} h_k z^{-k} \tag{1}$$

of McMillan degree n be given. The *positive realization problem* for $H(z)$ is to find, for some dimension $N \geq n$, a matrix $A \in \mathbb{R}_+^{N \times N}$ and column vectors $b, c \in \mathbb{R}_+^N$ with nonnegative entries such that $H(z) = c^T(zI - A)^{-1}b$ holds. The *minimality problem* is to find the smallest possible value of N. The nonnegativity of the entries of A, b, c, reflecting *physical constraints*, is a natural restriction in many applications. For a recent overview of the theory of positive systems and its applications we refer to [6].

A complete answer to the "realization problem" (i.e. decide, in finite steps, whether $H(z)$ admits positive realizations and, if so, construct one) has been presented in [1] and [7] (see also [10] where a finite algorithm for checking the nonnegativity of the impulse response sequence of $H(z)$ is given). However, much less progress has so far been made with regard to the "minimality problem". It is a basic result in linear system theory that, without the nonnegativity restriction, canonical realizations of $H(z)$ of dimension n always exist. However, it is well-known that the nonnegativity constraint may force N to be strictly greater than n. Moreover, it seems to be very difficult to give tight lower and upper bounds on the dimension N of positive realizations of $H(z)$. A useful estimate from above is given in [8] for primitive transfer functions with real simple poles. A general estimate from below is also presented in [8]. In [2] the authors give an interesting example of how the nonnegativity constraint may force realizations of large dimension even in the case of transfer functions of McMillan degree 3 with positive simple poles. For the class of third order transfer functions with positive simple poles the minimality problem was recently tackled in [3], where the important case $N = 3$ is characterized.

In the course of this brief we use a graph-theoretical approach to provide a non-trivial lower bound on the value of N for a special class of transfer functions $H(z)$. In particular, we *assume that the impulse response sequence* $(h_k)_{k\geq 1}$ *of* $H(z)$ *is nonnegative and there exists an instant* k_0 *such that* $h_{k_0} = 0$ *and* $h_k > 0$ *for all* $k > k_0$. Transfer functions with this property may be regarded as 'extremal' cases in positive system theory since $h_{k_0} = 0$ is the smallest admissible value of the impulse response at any time instant k_0. The graph-theoretical approach for positive system realizations has already proved to be useful in revealing reachability and controllability properties of positive linear systems (see [4] and [5]). However, this approach is new in proving dimension estimates, and provides a new insight into how the nonnegativity constraints may force realizations of large dimensions.

The paper is organized as follows. In Section II we give some standard definitions and a preliminary result in digraph theory which will be needed later in the proof of Theorem 1. Section III contains the main result of the paper: a general lower bound on the dimension N is presented in terms of k_0. Finally, we give examples illustrating the results.

II. Notation and preliminary results

In this section we recall some standard digraph terminology and prove an auxiliary result on digraphs in Lemma 1.

Let $A \in \mathbb{R}_+^{N\times N}$ denote an arbitrary nonnegative matrix. The digraph $\mathcal{G}$ corresponding to A is defined as follows. $\mathcal{G}$ is a digraph on N vertices $v_1, v_2, \ldots v_N$ and there is an arc (v_j, v_i) in $\mathcal{G}$ if and only if $a_{i,j} > 0$.

Let v_i, v_j be two (possibly identical) vertices of $\mathcal{G}$. We will use the following terminology (see, for example [5]):

a walk of length $l > 0$ from v_i to v_j is any sequence $(v_i = v_{i_1}, v_{i_2}, \ldots, \ldots, v_{i_{l+1}} = v_j)$ of vertices such that there exist arcs from v_{i_s} to $v_{i_{s+1}}$ for all $1 \leq s \leq l$. A walk of length 0 between v_i and v_j will simply mean that $v_i = v_j$ and we consider the vertex v_i on its own as a degenerate walk.

a path of length $l > 0$ from v_i to v_j is a walk of length l where all the vertices $v_i = v_{i_1}, v_{i_2}, \ldots, v_{i_{l+1}} = v_j$ are distinct. A path of length 0 is the same as a walk of length 0.

a cycle of length $l > 0$ from v_i is a closed walk $v_i = v_{i_1}, v_{i_2}, \ldots, v_{i_{l+1}} = v_i$ such that all vertices are distinct, except for $v_{i_1} = v_{i_{l+1}}$. Cycles of length 0 are not defined.

We also introduce the following simple *operations* on the digraph $\mathcal{G}$:

adding a cycle to a walk means that if $(v_i = v_{i_1}, v_{i_2}, \ldots, v_{i_{l+1}} = v_j)$ is a walk between v_i and v_j, and $(v_{i_s} = v_{j_1}, v_{j_2}, \ldots, v_{j_t} = v_{i_s})$ is a cycle from v_{i_s} (for some $1 \leq s \leq l+1$) then we consider the new walk $(v_i = v_{i_1}, v_{i_2}, \ldots, v_{i_s} = v_{j_1}, v_{j_2}, \ldots, v_{j_t} = v_{i_s}, v_{i_{s+1}}, \ldots, v_{i_{l+1}} = v_j)$.

deleting a cycle from a walk means that if $(v_i = v_{i_1}, v_{i_2}, \ldots v_{i_s} = v_{j_1}, v_{j_2}, \ldots, v_{j_t} = v_{i_s}, v_{i_{s+1}}, \ldots, v_{i_{l+1}} = v_j)$ is a walk containing a cycle from v_{i_s} then we consider the new walk $(v_i = v_{i_1}, v_{i_2}, \ldots v_{i_s}, v_{i_{s+1}}, \ldots, v_{i_{l+1}} = v_j)$.

descending to a path P from a walk R means that we delete some (possibly none) cycles from R and we obtain P.

Note that from any walk between v_i and v_j we can descend to a path (possibly of length 0) between v_i and v_j by deleting cycles one after the other.

The results on digraphs in this section will be needed in the proof of Theorem 1.

Given a digraph $\mathcal{G}$, for a given pair of (possibly identical) vertices (v_i, v_j) we would like to *characterize the numbers l such that there exists a walk of length l between v_i and v_j.*

Take *a path P_1 of length l_1* between v_i and v_j. (If there exist no paths from v_i to v_j, then there exist no walks either.) Consider all cycles from all vertices of the path P_1. If among the cycles there exist some, which contain some "new" vertices (not yet included in the path P_1), then choose any one of these cycles and add it to the path P_1, thus obtaining a new walk R_2. Now apply the same procedure to R_2: consider all cycles from all vertices of the walk R_2, and if there exist some which contain some "new" vertices (not yet included in the walk R_2), then choose any one of these cycles and add it to the walk R_2, thus obtaining a new walk R_3. Keep repeating this procedure until you arrive at *a maximal walk extension* $W_0 := (v_i = v_{k_1}, v_{k_2}, \ldots, v_{k_{w_0+1}} = v_j)$ *of length* w_0 such that the application of the procedure to W_0 yields no further new vertices. It is clear that such a walk W_0 is obtained after a finite number of applications of the described procedure. It is easy to see that W_0 contains, besides the vertices contained in P_1, all the vertices of $\mathcal{G}$ "reachable" from P_1 by addition of cycles. Note also that W_0 is not uniquely determined by P_1 (but the *vertices* contained in W_0 *are*).

Now, consider all cycles $C_1, C_2, \dots C_Q$ from all vertices of W_0. Let $c_1, c_2, \dots, c_q$ denote the pairwise different lengths appearing among the lengths of these cycles. *Let d denote* the greatest common divisor of the numbers $c_1, c_2, \dots, c_q$. (If $q = 0$, then d is not defined. In this case it is clear that the only walk (from v_i to v_j), from which we can descend to P_1 is, in fact, P_1 itself.) In the terminology of the preceding paragraphs we have the following

Lemma 1: Assume that for a specified path P_1 of length l_1 and some of its maximal walk extensions W_0 of length w_0 (as above) we have $q > 0$, so that d is defined. For any positive integer $l \geq \frac{N(N+1)}{2} - 1 + N^2$ the following are equivalent:

(i) there exists a walk R of length l between v_i and v_j from which we can descend to P_1,

(ii) $l \equiv w_0 \equiv l_1 \pmod{d}$.

Proof. Assume (i). It is clear that $w_0 \equiv l_1 \pmod{d}$, because W_0 can be obtained from P_1 by adding some of the cycles C_k. Assume R is an arbitrary walk of length l (between v_i and v_j), from which we can descend to P_1. This means that we can delete some cycles C_k from R and obtain P_1. Therefore $l \equiv l_1 \pmod{d}$.

Assume now (ii). First we show that $w_0 \leq \frac{N(N+1)}{2} - 1$ (recall that N denotes the number of vertices in $\mathcal{G}$). We use mathematical induction with respect to the number m of applications of the described procedure. Let R_m denote the walk obtained after m applications of the procedure. Let l_m and s_m denote the length of R_m and the number of different vertices contained in R_m, respectively. We prove that $l_m < \frac{s_m(s_m+1)}{2}$. Initially (i.e. for $m = 0$), we have the path P_1 of length l_1, containing $l_1 + 1$ different vertices. Clearly $l_1 < \frac{(l_1+1)(l_1+2)}{2}$ holds. We make the inductive assumption that $l_m < \frac{s_m(s_m+1)}{2}$. We apply the procedure again, and add a new cycle C_{m+1} (containing some new vertices) to R_m. The length c_{m+1} of the cycle C_{m+1} is not greater than s_{m+1}. Therefore $l_{m+1} = l_m + c_{m+1} < \frac{s_m(s_m+1)}{2} + s_{m+1} \leq \frac{s_{m+1}(s_{m+1}+1)}{2}$, where the last inequality follows from the fact that $s_{m+1} \geq s_m + 1$. This completes the induction.

Now, considering that the number of vertices contained in W_0 cannot be greater than N, we get $w_0 < \frac{N(N+1)}{2}$, as desired.

The next step is to prove that there exists a positive integer L_0 such that every integer

l divisible by d and not smaller than L_0 can be decomposed as $l = \sum_{j=1}^{q} \alpha_j c_j$ where all the coefficients α_j are nonnegative integers.

We can assume that $c_1 < c_2 < \cdots < c_q$. Assume also, for the moment, that $d = 1$. We prove by induction with respect to q that $L_0 \le c_q^2$. If $q = 1$, then $c_1 = d = 1$ by assumption, and the statement is trivial. Assume the statement is true for an integer $q \ge 1$, and take numbers $c_1 < c_2 < \cdots < c_{q+1}$ with greatest common divisor $d = 1$. Denote by d_q the greatest common divisor of the numbers $c_1, c_2, \ldots, c_q$. Then, by applying the inductive assumption to the numbers $\frac{c_1}{d_q}, \frac{c_2}{d_q}, \ldots, \frac{c_q}{d_q}$, we see that any number l divisible by d_q and $l \ge \frac{{c_q}^2}{d_q}$ can be decomposed as $l = \sum_{j=1}^{q} \alpha_j c_j$ where all the integers α_j are nonnegative. Now assume that $l \ge \frac{c_q^2}{d_q} + (d_q - 1)c_{q+1}$. Then, whatever the remainder of l by d_q is, we can deduct some multiple of c_{q+1} so that $l - c \cdot c_{q+1}$ is divisible by d_q (with $0 \le c \le d_q - 1$). Therefore, l can be decomposed as $l = c \cdot c_{q+1} + \sum_{j=1}^{q} \alpha_j c_j$, as desired. In order to complete the argument of the case $d = 1$, it is enough to notice that $\frac{c_q^2}{d_q} + (d_q - 1)c_{q+1} \le c_{q+1}^2$.

The case $d > 1$ follows easily. Indeed, if $d > 1$, then consider the numbers $\frac{c_1}{d}, \frac{c_2}{d}, \ldots, \frac{c_q}{d}$ with greatest common divisor 1, and apply the result above. We obtain that for an integer l the conditions $d|l$ and $\frac{l}{d} \ge \frac{c_q^2}{d^2}$ imply $\frac{l}{d} = \sum_{j=1}^{q} \alpha_j \frac{c_j}{d}$. Hence $L_0 \le \dfrac{c_q^2}{d^2} d \le c_q^2$.

Considering that $c_q \le N$ we deduce that $L_0 \le N^2$.

Summarizing the results we obtain that the conditions $l \ge \frac{N(N+1)}{2} - 1 + N^2 \ge w_0 + L_0$ and $l \equiv w_0 \pmod{d}$ imply that l can be decomposed as $l = w_0 + \sum_{j=1}^{q} \alpha_j c_j$. Accordingly, we can add some cycles C_j to the walk W_0 and obtain a walk of length l from v_i to v_j.

III. A lower bound on N

In this section we provide a general lower bound on N in terms of k_0 for the class of transfer functions specified in the Introduction. Namely, we *assume that every* $h_k \ge 0$, *there exists an index* $k_0 \ge 1$ *such that* $h_{k_0} = 0$, *and* $h_k > 0$ *for all* $k > k_0$.

Theorem 1: With notation as above, the dimension N of any positive realization of $H(z)$ satisfies $\frac{N(N+1)}{2} - 1 + N^2 \ge k_0$.

Proof. Assume that $c = (c_i)_{i=1}^{N}$, $A = (a_{i,j})_{i,j=1}^{N}$, $b = (b_i)_{i=1}^{N}$ is a positive realization of $H(z)$ in N dimensions. We will use the notation $A^k := (a_{i,j}^{(k)})_{i,j=1}^{N}$ (for $k = 0, 1, 2, \ldots$). $\mathcal{G}$ will denote the digraph corresponding to A. By assumption, we have $c^T A^{k_0 - 1} b = 0$ and $c^T A^{k-1} b > 0$ for all $k > k_0$. This means that for all subscripts i_1, i_2 we have $c_{i_2} a_{i_2,i_1}^{(k_0-1)} b_{i_1} =$

0. Furthermore, for all $k > k_0$ we can find indices $i_{1,k}$, $i_{2,k}$ such that $c_{i_{2,k}} a^{(k-1)}_{i_{2,k},i_{1,k}} b_{i_{1,k}} > 0$. In the graph $\mathcal{G}$, a vertex v_i will be called a *vertex of input* if $b_i > 0$, and a *vertex of output* if $c_i > 0$. (Note that a vertex can be both a vertex of input and output at the same time.) The assumptions above mean that for all $k > k_0$, vertex $v_{i_{1,k}}$ is a vertex of input, $v_{i_{2,k}}$ is a vertex of output, and there exists a walk of length $k-1$ from $v_{i_{1,k}}$ to $v_{i_{2,k}}$ in $\mathcal{G}$.

We have seen in Lemma 1 that for a fixed path P_1 (between some fixed vertices v_i and v_j), the numbers l with the properties that $l \geq \frac{N(N+1)}{2} - 1 + N^2$ *and* there exists a walk of length l between v_i and v_j from which it is possible to descend to P_1, are characterized by a certain equivalence class of integers modulo a number d. (Or, possibly, there exist no such numbers l at all, which can happen if $q = 0$ and d is not defined.) Consider now all the possible paths $(P_j)_{j=1}^S$ from v_i to v_j. It is clear that the numbers l with the properties that $l \geq \frac{N(N+1)}{2} - 1 + N^2$ *and* there exists a walk of length l between v_i and v_j, are characterized by the union of equivalence classes corresponding to the paths $(P_j)_{j=1}^S$. Therefore, the numbers l such that $l \geq \frac{N(N+1)}{2} - 1 + N^2$ *and* there exists a walk of length l between *some* vertex of input and *some* vertex of output of $\mathcal{G}$, are also characterized by the union of certain equivalence classes. Now, there are two possibilities: either the union of these equivalence classes contains *each* number $l \geq \frac{N(N+1)}{2} - 1 + N^2$ or it does *not*. In the first case we have $k_0 - 1 \leq \frac{N(N+1)}{2} - 1 + N^2 - 1$. In the second case we conclude that there are infinitely many numbers not covered by the union of the equivalence classes. (Indeed, if l is not covered by the union of the equivalence classes, then $l + \prod_{j=1}^S d_j$ is not covered either, where d_j denotes the modulus of the equivalence class corresponding to P_j.) Therefore, in this case, there would be infinitely many 0's contained in the impulse response sequence of $H(z)$, contradicting the assumptions of the theorem. Hence, under our assumptions, only the first case is possible, i.e. $k_0 \leq \frac{N(N+1)}{2} - 1 + N^2$.

Remark 1. With a refinement of the digraph arguments above the lower bound on N may be possible to improve slightly. In fact, $k_0 \leq N^2 - 2N + 2$ seems possible to achieve (cf. Example 2 below). For small values of N this can be proved by checking all configurations of digraphs on N vertices, and determining the greatest possible value of k_0.

Finally, we give two examples illustrating the results above.

Example 1. Take the transfer function

$$H(z) = \frac{1}{z-1} + \frac{0.75\cdot(\frac{20}{19})^{25}}{z+0.95} - \frac{0.75\cdot(\frac{20}{19})^{25}}{z-0.95} + \frac{0.5\cdot(\frac{10}{9})^{25}}{z-0.9} (\approx \frac{1}{z-1} + \frac{2.7037786}{z+0.95} - \frac{2.7037786}{z-0.95} + \frac{6.9647778}{z-0.9}).$$

It is not hard to check that the impulse response of $H(z)$ is nonnegative, and $h_{26} = 0$ and $h_k > 0$ for all $k > 26$. Although $H(z)$ is a transfer function with four real simple poles, we can conclude from Theorem 1 that the dimension N of any positive realization of $H(z)$ must be at least 5. In fact, by checking all digraphs on 5 vertices it is not hard to prove that $N \geq 6$ must hold. Having established this lower bound, it would be interesting to know what the actual minimal value of N is.

Consider now the sequence of transfer functions

$$H_m(z) = \frac{1}{z-1} + \frac{0.75\cdot(\frac{20}{19})^{\frac{m(m+1)}{2}+m^2-1}}{z+0.95} - \frac{0.75\cdot(\frac{20}{19})^{\frac{m(m+1)}{2}+m^2-1}}{z-0.95} + \frac{0.5\cdot(\frac{10}{9})^{\frac{m(m+1)}{2}+m^2-1}}{z-0.9}$$

(for $m = 4, 5, 6, \ldots$). Each function $H_m(z)$ satisfies the conditions of Theorem 1 with $k_0 = \frac{m(m+1)}{2} + m^2$. Although $H_m(z)$ is a transfer function with four fixed (real) simple poles, it cannot have a positive realization of dimension less than $m + 1$, by Theorem 1. In [2] a similar example was presented by using convex cone analysis. Although the dimension estimate given in [2] is tighter, an advantage of Theorem 1 above is that it can be applied to a wider class of transfer functions.

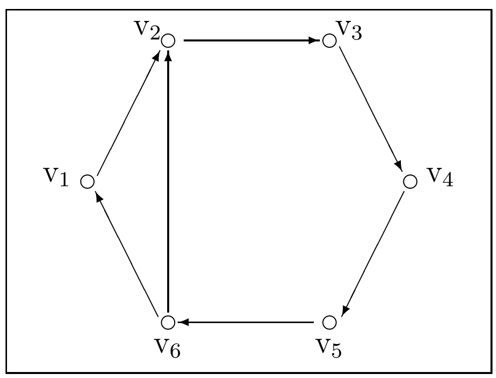

Fig. 1. The digraph corresponding to the matrix A for $N = 6$ in Example 2.

Example 2. The following example shows that the *order* of the bound presented in Theorem 1 cannot be improved by using digraph techniques only. (For an improved bound it seems necessary to consider the location of the poles of $H(z)$.)

Define the matrix $A \in \mathbb{R}_+^{N\times N}$, such that $a_{i+1,i} = 1$ (for $i = 1, 2, \ldots, N-1$), $a_{1,N} = a_{2,N} = 1$, and all other entries are 0. (Figure 1 shows the digraph corresponding to A) Define $b := (1, 0, \ldots, 0)^T$ and $c := (1, 0, \ldots, 0)^T$. Then the triple (c, A, b) gives a positive realization of the transfer function $H(z) := c^T(zI - A)^{-1}b$ in N dimensions. Furthermore, it is not hard to check that, for the impulse response sequence corresponding to $H(z)$, we have $h_{N^2-2N+2} = 0$ and $h_k > 0$ for all $k > N^2 - 2N + 2$. This shows that the estimate of Theorem 1 cannot be improved beyond $k_0 \leq N^2 - 2N + 2$.

IV. Conclusions

In the course of this paper we used a digraph approach for a special class of transfer functions to present a lower bound on the dimension of positive realizations. The transfer functions under consideration were specified by the assumption that there exists a time instant k_0 at which the impulse response of $H(z)$ is 0 but the (always nonnegative) impulse response becomes strictly positive for all $k > k_0$. Transfer functions with this property can be regarded as extremal cases in positive system theory, since $h_{k_0} = 0$ is the smallest admissible value of the impulse response at any time instant k_0.

References

[1] B.D.O. Anderson, M. Deistler, L. Farina, and L. Benvenuti, "Nonnegative realization of a linear system with nonnegative impulse response," *IEEE Trans. Circuits Syst. I*, vol. 43, no.2, pp 134-142, Feb. 1996.

[2] L. Benvenuti, L. Farina, "An example of how positivity may force realizations of 'large' dimensions," *Syst. Contr. Lett.*, vol. 36, no.4, pp 261-266, 1999.

[3] L. Benvenuti, L. Farina, B.D.O. Anderson and F. De Bruyne "Minimal positive realizations of transfer functions with positive real poles," *IEEE Trans. Circuits Syst. I*, vol. 47, no.9, pp 1370-1377, Sept. 2000.

[4] R. Bru, L. Caccetta, S. Romero, V. Rumchev and E. Sanchez, "Recent developments in reachability and controllability of positive linear systems," *Proc. 15th IFAC World Congress*, 2002, July 21-26, Barcelona, Spain.

[5] L. Caccetta and V. Rumchev, "A survey of reachability and controllability for positive linear systems," *Annals of Operations Research*, vol. 98, pp 101-122, 2000.

[6] L. Farina and S. Rinaldi, *Positive linear systems: Theory and applications.* New York: Wiley, 2000.

[7] L. Farina, "On the existence of a positive realization," *Syst. Contr. Lett.*, vol. 28, no.4, pp 219-226, 1996.

[8] C. Hadjicostis, "Bounds on the size of minimal nonnegative realizations for discrete-time LTI systems," *Syst. Contr. Lett.*, vol. 37, no.1, pp 39-43, 1999.

[9] Y. Ohta, H. Maeda, and S. Kodama, "Reachability, observability and realizability of continuous-time positive systems," *SIAM J.Control Optim.*, vol. 22, no.2, pp. 171-180, 1984.

[10] B. Nagy, M. Matolcsi, "Algorithm for positive realization of transfer functions," (preprint)