On the use of quasi-positive versus positive state-space models of externally positive discrete-time systems

R. Guidorzi, G. Marro, R. Diversi and U. Soverini

Abstract—Positive state-space models describe large classes of processes in econometry, epidemiology, biology, ecology, chemistry, hydraulics and logistics. These models satisfy strict algebraic conditions that can be directly fulfilled when the models are obtained by means of traditional modeling techniques by selecting the state variables in a “natural” way i.e. by associating a well-defined physical meaning to every variable. The situation is more critical when positive state-space models must be obtained by means of realization techniques from transfer functions since, in this case, the fulfillment of positivity conditions could call for the introduction of spurious dynamics and non minimal parameterizations. A possible alternative consists in using quasi-positive models; this paper discusses the pros and cons of these solutions.

I. INTRODUCTION

Externally positive systems are defined as systems whose impulse response is nonnegative and that, consequently, generate nonnegative responses when driven by nonnegative input sequences. State-space models are defined as positive when any state-space trajectory generated by a nonnegative input sequence from a nonnegative initial state evolves inside the positive orthant and is mapped into a positive response. This property implies that all entries of the input and output distribution matrices are nonnegative; moreover the impulse response is nonnegative so that positive state-space systems are also externally positive.

Positive systems play a role of paramount importance in the real world since many relevant processes are intrinsically positive; consider, for instance, the family of hydraulic systems, the diffusion of pollutants in the environment and drug absorption to mention only few well-known contexts. As a consequence, the role of positive models in control, observation, diagnosis, filtering and other problems is relevant and, since most mathematical tools presently available to deal with these problems concern state-space models, an important area of research is focused on the analysis of the mathematical properties of state-space positive models and on their realization i.e. the problem of deducing models of this kind from external descriptions.

This research, that has seen the pioneer contributions of Luenberger [1], Anderson [2], Farina and Rinaldi [3], Farina and Benvenuti [4] and many others [5], [6], [7], [8], [9], [10], has outlined that, despite the simple and natural definition of state-space positivity, the associated mathematical properties are somehow complex and, sometimes, difficult to fulfill. Part of these problems is due to the fact that while external positivity is an intrinsic invariant property of a model, possibly deriving from physical properties, the links of these properties with state-space positivity are much more elusive and not invariant with respect to changes of basis in the state-space; positive realizations can require orders larger than that of the corresponding transfer function or even not exist [4]. This situation derives from the fact that the conditions for state-space positivity exceed the requirements necessary for external positivity. To overcome these problems, the more relaxed definition of quasi-positivity has been introduced in [11]. Quasi-positive state-space models limit the positivity conditions to the state-space trajectories actually possible for the considered system and are endowed with the following properties: 1) they exist for every externally positive transfer function; 2) they do not require any order increase with respect to the transfer function one; 3) they can always be associated with a minimal parameterization.

This paper proposes some comparisons between the use of positive and quasi-positive realizations of the same transfer function and shows that the additional dynamics that must be added, in some cases, to fulfill positivity conditions do not play any role in control and filtering applications making the alternative use of quasi-positive models in these applications advantageous. The considerations are limited, like in [11], [4] and in most other works concerning positive realization, to the case of single-input single-output (SISO) systems; multi-input multi-output (MIMO) quasi-positive realization will be treated in a future paper.

The contents are organized as follows. Section 2 reports some considerations on the advantages and possible problems associated with the use of state-space positive models and compares the practical use of these models with that of quasi-positive ones. The use of positive and quasi-positive models in Errors-in-Variables (EIV) and Kalman filtering examples is then described in Sections 3 and 4. Some concluding remarks are eventually reported in Section 5.

II. POSITIVE VERSUS QUASI-POSITIVE MODELS

Systems whose “natural” state is intrinsically constrained in the positive orthant are frequent in econometry, epidemiology, biology, ecology, chemistry, hydraulics, logistics and also in specific applications in other fields. This feature has not been explicitly exploited for many years where positive systems have been treated as generic linear systems with additional constraints introduced for inducing the desired properties in the system behavior. More recently some
specific properties, like reachability and controllability, that exhibit substantial differences from those of non positive systems, have been studied with reference to state-space representations and this, in turn, has spurred remarkable interest on the positive realization problem. When the use of positive state-space models to describe externally positive systems is suggested or requested by specific problems, the question to be solved concerns how obtaining models of this kind possibly endowed with some additional useful features like minimality.

A first context can be associated with the cases in which the model is obtained by means of classical modeling techniques and this requires a complete knowledge of the physical structure of the system to be modeled and of the physical laws (if any) describing its behavior. This is by far the most favorable situation because, by properly selecting the state variables (usually on the basis of their physical meaning) positive models can be directly obtained.

A second context concerns situations where the initial information consists in the system transfer function. This happens, for instance, when the model has been obtained by means of identification techniques from input/output measures performed on the process to be modeled. Once that the external positivity has been verified (and this could require a large number of checks, conceptually infinite, on the impulse response samples), some problems could be approached by using only the transfer function while others could call for the use of a state-space realization.

A first possibility concerns the use of positive state-space models that satisfy the following set of conditions.

Given an \((n \times n)\) nonnegative matrix \(M\), \(\sigma (M)\) will denote the spectrum of \(M\), i.e. \(\sigma( M ) = \text{eig}( M ) = \{ \lambda_1, \ldots, \lambda_n \}\). The spectral radius of \(M\) is defined as \(\rho( M ) = \max_i | \lambda_i |, \) \((i = 1, \ldots, n)\). Every eigenvalue with maximal modulus is defined as dominant eigenvalue, \(\lambda_d\); thus \(\rho( M ) = | \lambda_d |\). The triple \((A, b, c)\) defining the positive discrete-time single-input single-output (SISO) state-space model

\[
x(k + 1) = Ax(k) + bu(k),
\]

\[
y(k) = cx(k)
\]

satisfies the following conditions \([1], [2], [3], [5], [12], [13], [14], [15]\):

1) The matrices \((A, b, c)\) have positive or null entries.
2) The impulse response \(w(k) = cA^{k-1}b\) is positive or null for every \(k > 0\).
3) The set of reachable states of system \((1)-(2)\) is the reachability cone

\[
\mathcal{R}_c = \text{cone} \{ b, Ab, A^2b, \ldots \}.
\]

4) Perron–Frobenius theorem. The dominant eigenvalues of \(A\) are all the roots of \(\lambda^k - \rho(A)^k = 0\) for some (also more than one) values of \(k = 1, 2, \ldots, n\). One of the dominant eigenvalues is positive real, i.e. \(\rho(A) \in \sigma(A)\) and, for every dominant eigenvalue, \(\lambda_d, \deg \rho(A) > \deg \lambda_d\).
5) Karpelevich theorem. The regions of the complex plane that contain the eigenvalues of a matrix \(A\) having spectral radius \(\rho\) are symmetric with respect to the real axis, are included in the disc \(|z| \leq 1\), and intersect the circle \(|z| = 1\) in the points \(e^{2\pi i a/b}\) where \(a\) and \(b\) run over the relatively prime integers satisfying the condition \(0 \leq a \leq b \leq n\). The boundary of these regions consists of these points and of curvilinear arcs connecting them in circular order.

The previous conditions are not necessarily met by a generic realization of a given positive transfer function \(G(z)\) and their fulfillment could be impossible by means of a simple change of basis in the state-space. Moreover, in some cases a positive realization could require an order larger than that of \(G(z)\) or even be impossible with finite-order models \([15]\). Another critical aspect concerns the possible presence of some eigenvalues of \(A\), not present as poles of \(G(z)\), with negative real part (see, for instance Examples 7 and 8 in \([15]\)); this would mean that no continuous positive model exists for the considered process.

A second possibility consists in using quasi-positive systems, defined as follows \([11]\).

**Definition 1:** A Quasi-Positive SISO system consists in a triple \((A, b, c)\) describing a model of the type \((1)-(2)\) that satisfies the following conditions:

1) Every state trajectory generated by a reachable initial state by applying a nonnegative control sequence \(u(\cdot)\) assumes positive or null values.
2) Every output trajectory generated by a reachable initial state by applying a nonnegative control sequence \(u(\cdot)\) assumes positive or null values.

The properties of quasi-positive systems are the following:

a) The impulse response \(w(k) = cA^{k-1}b\) is positive or null for every \(k > 0\).

b) All states reachable from \(x(0) = 0\) for \(u(\cdot) \geq 0\) are positive or null.

c) The dynamic matrix \(A\) transforms any reachable state \(x \in \mathcal{R}_c\) into a positive or null state.

d) The entries of \(b\) and \(c\) are positive or null.

e) The cone of the reachable states for \(u(\cdot) \geq 0\), \(\mathcal{R}_c = \text{cone} \{ b, Ab, A^2b, \ldots \}\), is included in the positive orthant of \(\mathbb{R}^n\).

It can be observed that, differently from positive systems, the entries of \(A\) can assume also negative values so that the state space \(X\) can contain positive non reachable states that are not transformed by \(A\) into positive states.

The relaxation in the positivity conditions introduced by quasi-positive systems makes it possible to apply, in the SISO case, simple and well-known procedures, like those recalled in \([11]\), that assure, for every externally positive system, the existence of a quasi-positive realization with the same order as the transfer function \(G(z)\) to be realized.

**Remark 1:** From a practical point of view, the use of a quasi-positive model instead of a positive one maintains the advantage of having positive state trajectories. Of course trajectories originated by non reachable initial states artificially introduced in the model will not necessarily respect the

1512
positivity conditions; if the model is asymptotically stable, however, the perturbed trajectories will converge to (positive) admissible ones.

As an example, consider the externally positive system described by the transfer function

$$G(z) = \frac{0.1372 z^2 + 0.0117 z - 0.1040}{z^3 - 2.5794 z^2 + 2.2157 z - 0.6338}$$

whose impulse response is reported in Figure 1.

A quasi-positive realization of (4) is given by the triple

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.6338 & -2.2157 & 2.5794 \end{bmatrix}, \quad b = \begin{bmatrix} 0.1372 \\ 0.3657 \\ 0.5351 \end{bmatrix}$$

$$c = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$  

The eigenvalues of $A$ are $\lambda_1 = 0.8076$, $\lambda_2 = 0.8808$ and $\lambda_3 = 0.8910$. The reachability cone, $\mathcal{R}_c = \text{cone} \{ b, Ab, A^2b, \ldots \}$, is reported in Figure 2 that shows also that the trajectory associated with the non reachable initial state $x(0) = [0 \ 1 \ 1]^T \notin \mathcal{R}_c$ for a given nonnegative input sequence $u(t)$ does not remain, during the initial transient, inside the positive orthant but returns into it within 20 transitions and then tends to $\mathcal{R}_c$. The trajectory of the first state component is reported in blue in Figure 3 versus the trajectory associated with the initial state $x(0) = [0 \ 0 \ 0]^T$.

The difference between the whole state trajectories is reported in Figure 4 that shows how it approaches zero for $t \to \infty$ in that associated with the free motion $x(k) = A^{k-1}[0 \ 1 \ 1]^T$.

### III. A NUMERICAL EXAMPLE IN EIV FILTERING

In this section it is shown how quasi-positive models can be used in the solution of a problem calling for the use of state-space models (filtering) without any loss of information with respect to the use of positive models while preserving the positivity of state-space trajectories.

The considered process is described by the following transfer function [15]

$$G(z) = \frac{4 z^2 + 1.6 z - 0.56}{z^3 + 0.2 z^2 - 0.88 z - 0.32}.$$  

A positive state-space realization with minimal dimension is [15]

$$A_p = \begin{bmatrix} 0 & 0 & 0 & 0.2304 \\ 1 & 0 & 0 & 0.5696 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.2000 \end{bmatrix}, \quad b_p = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$c_p = \begin{bmatrix} 4 \\ 0.8 \\ 2.8 \\ 1.4240 \end{bmatrix},$$

while a quasi-positive minimal realization is

$$A_{qp} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.32 & 0.88 & -0.20 \end{bmatrix}, \quad b_{qp} = \begin{bmatrix} 4.0 \\ 0.8 \\ 2.8 \end{bmatrix},$$

$$c_{qp} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

Both models have been used to perform a Monte Carlo simulation of 100 runs where the noiseless input and output sequences (in blue in Figures 5 and 6) have been corrupted with mutually correlated white additive noises $n_u$ and $n_y$ with variances $\sigma_u^2 = 1.2e-4$, $\sigma_y^2 = 1.8$ and covariance $\sigma_{uy} = 0.014$. The noise-corrupted input and output sequences are reported in red in Figures 5 and 6.
The Errors-in-Variables (EIV) filtering procedure described in [16] has then been applied; the results of a typical run are reported in Figures 7 and 8 where the noiseless sequences (blue) are compared with the filtered ones (red). Note that the considered EIV context allows not only the filtering of the state and output like in standard Kalman filtering but also that of the input.

\begin{align*}
G(z) &= (0.5256 z^2 - 0.01971 z - 0.4890) \cdot 10^{-3} \\
&= z^3 - 2.8240 z^2 + 2.6566 z - 0.8324.
\end{align*}

Figure 9 that shows how the spurious dynamics introduced to satisfy the positivity conditions concerns only the reachable but not observable part of the model and, consequently, does not influence the transfer function.

\begin{equation}
G(z) = c_p (z I - A_p)^{-1} b_p = c_2 (z I - A_{22})^{-1} b_2
\end{equation}

and, of course, the eigenvalues of \( A_{22} \) are the poles of (7). In this case the Kalman decomposition is obtained by referring to a basis of the state space of the positive model \( B = [B_1 B_2] \) partitioned into a basis, \( B_1 \) of the reachable and not observable subspace and into a basis, \( B_2 \) of the reachable and observable subspace. The positive model will be described, in this basis, by matrices with the following structure

\begin{align*}
A'_p &= B^{-1} A_p B = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, & b'_p &= B^{-1} b_p = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},
\end{align*}

\begin{align*}
c'_p &= c_p B = \begin{bmatrix} 0 \\ c_2 \end{bmatrix};
\end{align*}

it must, however, be observed that the reachable and the observable subsystem will be quasi-positive only by selecting a suitable basis \( B_2 \).

**IV. A NUMERICAL EXAMPLE IN Kalman filtering**

The model considered in the following describes a well-known and important positive process, the diffusion of an orally assumed drug in the three compartments constituted by the stomach, the blood and the less perfused organs. The model assumes as state variables the drug concentration (\( \mu g / m \ell \)) in the blood (first compartment), in the stomach (second compartment) and in the less perfused organs (third compartment) and has been obtained by discretization, with a sampling interval of 1 minute, of the continuous model described in [17]; the input is the drug assumption in \( \mu g \).

The state space model, that assumes as output the drug concentration in the blood, is described by the matrices

\begin{equation}
A_p = \begin{bmatrix} 0.9049 & 0.01572 & 0.06134 \\ 0.01572 & 0.9836 & 0.0005227 \\ 0 & 0 & 0.9355 \end{bmatrix}, & b_p = \begin{bmatrix} 0.0005256 \\ 0.000002948 \\ 0.01612 \end{bmatrix},
\end{equation}

\begin{equation}
c_p = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}
\end{equation}

and the corresponding transfer function is

\begin{equation}
G(z) = \frac{(0.5256 z^2 - 0.01971 z - 0.4890) \cdot 10^{-3}}{z^3 - 2.8240 z^2 + 2.6566 z - 0.8324}.
\end{equation}
Figure 11 reports the state trajectories of the system (15)-(16) associated with the initial state \( x(0) = [20 \ 0 \ 0]^T \) assuming the oral drug assumption reported in Figure 10.

The same process has then been simulated by introducing state and observation noise with covariance matrices

\[
Q = \begin{bmatrix} 0.03415 & 0.001788 & 0.0001872 \\ 0.001788 & 0.03832 & 0.005056 \\ 0.0001872 & 0.005056 & 0.03481 \end{bmatrix}, \quad R = \begin{bmatrix} 0.6949 \end{bmatrix} \tag{18}
\]

The corresponding signal to noise ratios are 27, 21 and 31 dB for the trajectories and 24 dB for the output. In this condition the trajectories and the observed output are those reported in Figures 12 and 13.

\[
A_{qp} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.8324 & -2.6566 & 2.8240 \end{bmatrix}, \quad b_{qp} = \begin{bmatrix} 0.0005256 \\ 0.001464 \\ 0.002251 \end{bmatrix} \tag{19}
\]

\[
c_{qp} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{20}
\]

It can be immediately observed that, since the output distribution matrices \( c_p \) and \( c_{qp} \) are equal, the trajectories associated with the first state component in models (15)-(16) and (19)-(20) are also equal. This means that, if we are interested in the optimal filtering of this trajectory (or of the output), models (15)-(16) and (19)-(20) will give the same results.

The positive model (15)-(16), however, because of the physical meaning of all state variables should be, when available, preferred. The quasi-positive model (19)-(20) remains, however, an excellent choice when the model cannot be obtained by modeling techniques but only by means of identification procedures.

**Remark 3:** It can be observed that by using the (well-known) realization procedure outlined in [11], the output distribution matrix \( c_p \) is, in the SISO case, given by the first row of the identity matrix so that the first state variable coincides with the output and, consequently, is associated with a well-defined physical meaning. The same cannot be repeated, however, for the remaining state components.
V. CONCLUDING DISCUSSION AND REMARKS

Positive models constitute fascinating representations of many physical and abstract processes of paramount importance; the sophisticated mathematical theory that has been developed to describe the algebraic and geometrical properties of these models has outlined the severe constraints to be respected in their realization and the possibly critical aspects concerning their dimension and even their existence. Positive models can be constructed in various contexts like those that follow.

1) The model is obtained, by means of classical modeling techniques, on the basis of a complete knowledge of the process to be described and of the physical laws that describe its behavior. This is the ideal condition that leads in a natural way to positivity and to models whose order reflects the physical nature of the process (for instance number of compartments). In this case, of course, realization does not play any role.

2) The model is obtained, by realization, from the knowledge of the impulse response, \(G(z)\), or of the impulse response \(w(k)\). This situation could derive from the observation of the impulse response (as happens, for example, in some clinical tests like glucose tolerance test for diabetes or the bromsulphalein test for hepatic function). In some cases the whole procedure has to do more with identification than with realization or, to be more precise, with the realization of an identified response; the reasons are linked not only to measurement errors due to quantization and noise but also to the distributed-parameter nature of all real processes and to their nonlinearity. In situations of this kind a realization, i.e. a state space model, can be necessary because of the planned use of the model (e.g. optimal filtering or control).

3) The model is obtained by realizing an identified externally positive model (transfer function, polynomial model etc.). This case does differ from the previous one but, depending on the quality of the observations and on the conditions under which identification has been performed, the positivity constraints on the realization could be more difficult to fulfill.

In the last two situations the use of quasi-positive models is always possible and assures the positivity of all admissible state-space trajectories. Moreover the order of quasi-positive realizations equals the transfer function one and it is always possible to select a state space basis leading to a minimal parameterization.

The purpose of this paper has been to show, by means of simple numerical examples, that the use of quasi-positive models does not constitute a limit in the description of the positivity properties of a system; in fact the additive dynamics added in positive realization, when present, are reachable but not observable and do not reflect any physical property of the process.

REFERENCES