Contraction-based design of positive observers

Thach N. Dinh, Silvère Bonnabel, and Rodolphe Sepulchre

***Abstract*— We consider the problem of positive observer design for positive systems. We propose the design of a non-linear positive observer based on the use of generalized polar coordinates in the positive orthant. The contraction properties of the estimation error are studied thanks to the Hilbert projective metric. The optimization problem of finding the observer gains that maximize the contraction rate is addressed, and the simple two-dimensional case is discussed in detail.**

***Index Terms*— Positive systems, non-linear Luenberger-type positive observers, contraction properties, Hilbert metric.**

I. INTRODUCTION

Positive systems arise in several areas, where each coordinate of the state represents a positive quantity such as a population or a concentration, that does not have any physical meaning if it is not positive, see, for example, [7], [10], [5]. Positive observers aim at estimating the state variables from output measurements in such a way that the estimates are always positive, and thus admit a physical interpretation at all times, that is, even when the estimation error is not small. Positive observers for linear systems can be built under specific structural assumptions. In [9] and [6] structural properties, including observability, of positive systems have been explored, and observers for compartmental systems have been developed [11]. In [1] the positive observer design problem has been dealt with using coordinates transformations and the theory of positive realization [9], [2], thus generalizing the results in [11] and relaxing the conditions under which positive observers exist.

In a previous paper [3], the authors have advocated the use of generalized polar coordinates (that is, direction and norm) in solid cones on Banach spaces to build positive observers. The starting point of the approach is a theorem by G. Birkhoff which dates back to the 1950s, and which proves that a large class of positive mappings on cones admit contraction properties in the projective space for the so-called Hilbert metric. Thus, the idea underlying the method of [3] is that for those systems, a mere copy of the dynamics may yield an observer in the projective space, as the contraction properties of the dynamics imply that two arbitrary solutions (namely the true solution and the observer which, being a copy of the system, is also a solution) converge exponentially towards each other [8]. Once the direction of the true state is correctly estimated (i.e. the state is correctly estimated in the projective space), very mild assumptions on the output map allow to reconstruct its norm, and thus to estimate the whole state.

The present paper builds upon previous work [3] and focuses on the case of discrete-time positive time-varying linear systems in the orthant $\mathbb{R}^n_+$ in the standard form $x_{k+1} = A_k x_k + B_k u_k$ with output $y_k = C_k u_k$. The proposed Luenberger type candidate observer of the form $w_{k+1} = A_k w_k + B_k u_k + L_k(C_k w_k - y_k)$ contains correction terms in addition to the copy of the system dynamics in order to accelerate its *contraction rate*. The design is based on the use of Birkhoff theorem again, but here the goal is to reconstruct the linear estimation error $e_k = w_k - x_k$ in the projective space. Once the direction of the error e_k is correctly estimated, its norm can be estimated in turn. This latter step yields a non-linear estimated error $\hat{e}_k$ and the final positive observer is merely the sum $\hat{x}_k = w_k + \hat{e}_k$.

At each step, the contraction rate of the error system is directly linked to the so-called projective diameter of the matrix $A_k + L_k C_k$. The problem of finding the optimal gain matrix L_k that achieves the maximal contraction rate for the error system in the projective space is posed. It is non-convex and non-linear, but it is shown that simple heuristics suffice to determine a matrix gain L_k such that the contraction rate of $A_k + L_k C_k$ is higher than the one of A_k, implying that the observer converges faster that a mere copy of the dynamics. The design is in sharp contrast with the classical design of a Luenberger observer that focuses on the eigenvalues assignment (or spectral radius) of the linear error dynamics.

The paper is organized as follows: in Section II, the definition of positive linear observers and the Birkhoff-Bushell theorem are recalled. In Section III, a class of positive non-linear observers on the positive orthant is proposed. Section IV addresses the problem of gain tuning to enforce contraction. A numerical example is given in Section V. Concluding remarks are drawn in Section VI.

II. POSITIVE LINEAR SYSTEMS AND THE BIRKHOFF-BUSHELL THEOREM

A positive linear system is a linear system whose solutions live in the positive orthant at all times. Such systems arise naturally in several aeras where the state variables represent quantities that are intrinsically positive, such as concentrations or populations. For $x, y \in \mathbb{R}^n_+$, we denote $x \geq y$ when for all the components $x_i \geq y_i$. This operator is understood as a set of inequalities applied componentwise. Therefore,

T.N. Dinh and S. Bonnabel are with Centre de Robotique, MINES ParisTech, 75272 Paris, France thach.dinh@mines-paristech.fr, silvere.bonnabel@mines-paristech.fr

R. Sepulchre is with the Dept. of Electrical and Computer Engineering (Montefiore Institute, B28), University of Liège, 4000 Liège, Belgium, r.sepulchre@ulg.ac.be, and with INRIA Lille, F 59650 Villeneuve d'Ascq, France. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

a vector $x \in \mathbb{R}^n$ is said to be non-negative if $x \geq 0$ and positive if $x > 0$.

We define the orthant $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0, \forall i\}$ and $\mathbb{R}^n_{>0} = \{x \in \mathbb{R}^n : x_i > 0, \forall i\}$ to be its interior. $A \in \mathbb{R}^{n \times n}$ is said non-negative matrix if $A : \mathbb{R}^n_+ \to \mathbb{R}^n_+$, and if $A : \mathbb{R}^n_{>0} \to \mathbb{R}^n_{>0}$ we say that A is positive.

In this paper, we consider a linear time-varying system

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k \\ y_k &= C_k x_k \end{aligned} \tag{1}$$

where

- (i) A_k is a $n \times n$ positive matrix for all $k \geq 0$.
- (ii) B_k is a $n \times m$ non-negative matrix for all $k \geq 0$.
- (ii) C_k is a $p \times n$ non-negative matrix for all $k \geq 0$.

If the three conditions above are met, the linear system (1) is said to be positive in the following sense: if the initial state x_0 is non-negative and the control input u_k is non-negative for all $k \geq 0$, then the state x_k and the output y_k are non-negative at all time $k \geq 0$.

The Hilbert projective metric on $\mathbb{R}^n_{>0}$ is defined by:

$$d(x, y) = \max_{i,j} \log \left(\frac{x_i y_j}{x_j y_i} \right) \tag{2}$$

for any two vectors $x, y \in \mathbb{R}^n_{>0}$. It is understood as a projective metric in the sense that for all $\lambda, \mu > 0$ and x, y in the orthant we have $d(\lambda x, \mu y) = d(x, y)$. The projective diameter of a positive mapping A is defined as the diameter in the sense of the Hilbert metric of the image of the orthant by A, i.e.,

$$\Delta(A) = \sup\{d(Ax, Ay) | x, y \in \mathbb{R}^n_{>0}\} \tag{3}$$

and it may be infinite (take for instance the identity function). In the case of a positive linear mapping on $\mathbb{R}^n_{>0}$ the projective diameter can be expressed as [4]

$$\Delta(A) = \max \{\log [a_{ij} a_{kr} / a_{ir} a_{kj}] : 1 \leq i, j, k, r \leq n\} \tag{4}$$

where $A = (a_{ij})$ is the corresponding $n \times n$ matrix with positive entries. The latter formula can be understood in the following way. Note that, each column, say $a_{\bullet,j}$ for $1 \leq j \leq n$, of the matrix A represents the image of the j-th vector of the canonical basis of $\mathbb{R}^n$. But from (2), the distance between two vectors $x, y \in \mathbb{R}^n_{>0}$ can be written as $\max_{ij} \log \left(\frac{x_i y_j}{x_j y_i} \right)$. As the diameter of A is defined as the diameter of the image of the orthant, that is the diameter of the convex hull of the images $a_{\bullet,j}$'s of the base vectors, finding the projective diameter of A is equivalent to finding the maximal Hilbert metric between pairs of columns of A. The diameter is thus easily shown to be

$$\begin{aligned} \Delta(A) = \max_{j,r} d(a_{\bullet,j}, a_{\bullet,r}) &= \max_{j,r} \max_{i,k} \log \left(\frac{a_{ij} a_{kr}}{a_{kj} a_{ir}} \right) \\ &= \max_{i,j,k,r} \log \left(\frac{a_{ij} a_{kr}}{a_{kj} a_{ir}} \right) \end{aligned}$$

Note that, it is easily seen that the projective diameter will be finite if and only if all entries of A are positive.

The following theorem is at the core of our approach to the problem of positive observer design:

Theorem 1 (Birkhoff-1957): If A is a positive linear mapping we have for all $x, y \in \mathbb{R}^n_{>0}$

$$d(Ax, Ay) \leq \left(\tanh \left(\frac{\Delta(A)}{4} \right) \right) d(x, y) \tag{5}$$

As the hyperbolic tangent of any nonnegative number is always smaller than 1, we see that any positive linear mapping is a contraction. Moreover, if the projective diameter is finite, i.e. $\Delta(A) < \infty$, then A is a strict contraction. In particular, the Hilbert distance between two arbitrary solutions of the time-invariant system $x_{k+1} = A x_k$ tends exponentially to zero with rate $\gamma = \tanh \left(\frac{\Delta(A)}{4} \right) < 1$. This means, that for two arbitrary solutions $x_k, \tilde{x}_k$ we have $d(x_k, \tilde{x}_k) \leq \gamma^k d(x_0, \tilde{x}_0)$. Note that, the celebrated Perron-Froebenius theorem can be viewed as a consequence of the Birkhoff theorem as in the projective space A is proved to be a contraction. This implies using the Banach fixed point theorem any solution x_k will converge in direction to a vector, this vector being in fact an eigenvector of A. This is the reason why Birkhoff theorem is often considered as a generalization of the Perron-Froebenius theorem to general solid cones, and to homogeneous (possibly time-varying) non-linear positive maps [4]. As a final remark, note that, the contraction rate is monotonically linked to the diameter of the map A: a smaller diameter ensures a higher contraction rate.

III. A CLASS OF POSITIVE OBSERVERS FOR POSITIVE SYSTEMS

The reference [3] advocates a different observer design methodology for a class of positive linear systems on solid cones based on the Hilbert metric and Theorem 1. In the case where the considered solid cone is the positive orthant, those systems are of the form $x_{k+1} = A x_k$ where A is characterized by either a finite diameter, or A admits a power, say N, such that A^N has a finite diameter. For those systems, a mere copy of the system provides a simple positive observer which converges exponentially in the projective space. This is a straightforward consequence of Birkhoff's theorem. When the output is linear, it is then rather easy to reconstruct the norm of the true state as soon as the direction is sufficiently well estimated. Building upon this method, and focusing on the case of linear systems in the orthant, we propose a novel design methodology for positive observers allowing to embrace a broader class of time-varying systems of the form (1), where the observers include correction terms and gain matrices such that the the gains can be tuned to accelerate the convergence of the observer. As a result, the maps A_k (or their N-th power) need not necessarily have a finite projective diameter.

A. Non-linear Luenberger type positive observers for positive linear systems

For positive systems, a natural requirement is that the observers should provide state estimates that are also non-negative so they can be given a physical meaning at all times. In the present paper we are concerned with non-linear Luenberger-type positive observers for such systems.

Consider the system (1) in $\mathbb{R}^n_{>0}$. We propose to build a class of positive non-linear observers based on the following steps. First consider the Luenberger positive observer defined by

$$w_{k+1} = (A_k + L_k C_k) w_k + B_k u_k - L_k y_k \tag{6}$$

where w_k plays the role of a "surrogate" estimated state and where we assume $0 < w_0 < x_0$ (such an initial condition can always be found if we suppose that we have a lower bound on each coordinate of the state). The matrix gain $L_k \in \mathbb{R}^{n \times p}$ should satisfy the two following conditions:

(a) $A_k + L_k C_k$ is a $n \times n$ positive matrix having finite projective diameter for all $k \geq 0$.

(b) L_k satisfies the linear inequality $L_k y_k \leq B_k u_k$ for all $k \geq 0$.

Note that those two conditions are equivalent to the componentwise *linear* inequalities $L_k C_k > -A_k$ and $L_k y_k \leq B_k u_k$ at all times. Now consider the surrogate estimation error

$$e_k = x_k - w_k \tag{7}$$

For system (1) and observer (6), the error dynamics satisfy

$$e_{k+1} = (A_k + L_k C_k) e_k \text{ with } e_0 > 0 \tag{8}$$

The idea now is to use the Birkhoff theorem to reconstruct the direction of this surrogate error. Indeed consider the following positive system

$$\hat{e}_{k+1} = (A_k + L_k C_k) \hat{e}_k \text{ with } \hat{e}_0 > 0 \tag{9}$$

From the Birkhoff theorem we have $d(e_k, \hat{e}_k) \to 0$ exponentially as long as the diameter of $A_k + L_k C_k$ is bounded above by some fixed quantity $R > 0$. Thus, the normalized error $\hat{e}_k / \|\hat{e}_k\|$ will in this case provide a righteous estimation of the direction of the error. The final step consists of defining the state estimate as:

$$\hat{x}_k = w_k + \mu_k \frac{\hat{e}_k}{\|\hat{e}_k\|} \tag{10}$$

where for all $k \geq 0$ μ_k is defined by

$$\mu_k = \frac{\| y_k - C_k w_k \|}{\| C_k \hat{e}_k \|} \|\hat{e}_k\|$$

Note that, the quantity $\| y_k - C_k w_k \| / \| C_k \hat{e}_k \|$ can be replaced by the ratio of any component of the vector $y_k - C_k w_k$ with the same component of the vector $C_k \hat{e}_k$ (provided it is large enough to avoid problems in the division). This small modification can be advantageous when there is noise in the system and where it is not desirable to square the noise via a norm calculation.

The idea underlying the latter definition (10), is that, as soon as the gain matrix is well designed, $\hat{e}_k$ should allow to estimate the direction of the error e_k in the projective space, whereas μ_k allows to estimate its norm. The vector $\mu_k \frac{\hat{e}_k}{\|\hat{e}_k\|}$ is then a righteous estimation of the true error e_k.

The resulting observer is nonlinear and positive, as proved by the following result:

Proposition 1: For all $k \geq 0$, the estimated state defined by (10) satisfies $\hat{x}_k \geq 0$.

Proof: First, note that, because of assumptions (a) and (b), and the fact that $w_0 > 0$ initially, the estimate w_k defined by (6) is always positive. Then, because of the assumption that $w_0 < x_0$, and of assumption (a), we have that e_k is always positive. The estimate $\hat{x}_k$ defined by (10) is thus positive. ■

B. Convergence issues

The asymptotic convergence of the error e_k is characterized by the following proposition.

Proposition 2: Consider the system (1) and suppose there exists a non-positive $n \times p$ matrix L_k satisfying assumptions (a) and (b), and such that $\Delta(A_k + L_k C_k) \leq R$ for some $R > 0$, $\forall k \in \mathbb{N}$. Then the direction of error e_k is asymptotically well estimated as

$$d(\hat{e}_k, e_k) = d(\frac{\hat{e}_k}{\|\hat{e}_k\|}, e_k) \to 0 \quad \text{exponentially.}$$

where d is the Hilbert projective metric. Moreover, if there exist $\alpha, \beta, \epsilon > 0$ such that for all $k > 0$, $S(\frac{\hat{e}_k}{\|\hat{e}_k\|}, \epsilon) \subset \mathbb{R}^n_+$, where $S(z, \epsilon)$ is the ball of centre z and radius ϵ, and $\alpha \leq \|C_k\| \leq \beta$ then the norm of the error is asymptotically well estimated as

$$\left| \frac{\|e_k\|}{\mu_k} - 1 \right| \to 0 \quad \text{exponentially.}$$

Before proving the proposition, we discuss its interpretation. First, the theorem indicates convergence of the estimation error to zero in generalized polar coordinates. Indeed, the first part ensures that the direction of the error e_k is asymptotically reconstructed at an exponentially fast rate, and the second part that its norm is also asymptotically reconstructed exponentially fast, for a natural norm discrepancy based on the norms ratio. The norm ratio is a dimensionless quantity, and it is a meaningful way to measure discrepancy between norms, as unstable eigenvalues are often encountered in positive systems, and tends make norms of both the true and the estimated state go to infinity (see also [3] which advocates using this natural alternative error).

Secondly, if the ratio between the norms of the errors is sufficiently close to 1, the theorem implies that $d(\hat{x}_k, x_k)$ tends exponentially to zero. Indeed, we know from Theorem 3.5 in [4] that if $\|e_k\| = \mu_k$, then $d(\hat{x}_k, x_k) \leq d(\hat{e}_k, e_k)$. If the norm ratio is close to 1, the inequality may be violated, but in practice it is to be expected that $d(\hat{x}_k, x_k)$ be close to $d(\hat{e}_k, e_k)$, and thus exponentially convergent.

Finally, note that, the assumption of the proposition that $S(\frac{\hat{e}_k}{\|\hat{e}_k\|}, \epsilon) \subset \mathbb{R}^n_+$ essentially means that $\hat{e}_k$ does not converge to the boundary of the cone. It could have been replaced with the same condition on the true error e_k, but the choice is justified by the fact that the assumption as stated in the proposition can at all times be checked by the user, whereas the true error e_k is unknown. Note also that when the system is time-invariant, i.e., A, L, C are fixed, this latter assumption is automatically satisfied thanks to Birkhoff's theorem and the Banach fixed point theorem, which states then that there exits a positive vector v that is a fixed point of $A+LC$ in the projective space, and such that $d(e_k, v) \to 0$ exponentially.

Proof: The errors defined by (8) and (9) are solutions of the same positive system. Because of the assumption on the diameter of $A_k + L_k C_k$, and because of the Birkhoff-Bushell theorem, the considered system is contractive with a contraction ratio bounded away from 1, implying exponential convergence of the quantity $d(e_k, \hat{e}_k)$. As concerns the second result, note that

$$\mu_k = \frac{\|C_k e_k\|}{\|C_k \hat{e}_k\|}\|\hat{e}_k\|$$

letting $z_k = e_k/\|e_k\|$ and $\hat{z}_k = \hat{e}_k/\|\hat{e}_k\|$. Thus we have

$$\mu_k = \frac{\|C_k z_k\|}{\|C_k \hat{z}_k\|}\|e_k\|$$

Thus

$$\frac{\mu_k}{\|e_k\|} \leq \frac{1}{\alpha \cos\theta}|\|C_k z_k\| - \|C_k \hat{z}_k\|| \leq \frac{\beta}{\alpha \cos\theta}\|z_k - \hat{z}_k\|$$

where we have defined θ as the maximal angle between $\hat{z}_k$ and any line of the positive matrix C_k, whose cosine is bounded away from 0 as $\hat{z}_k$ has been assumed to be bounded away from the boundary of the cone. Using the fact that $z_k, \hat{z}_k$ have unit norm we can exploit a result of [4] which says that

$$\|z_k - \hat{z}_k\| \leq e^{d(z_k, \hat{z}_k)} - 1 = e^{d(e_k, \hat{e}_k)} - 1$$

Exponential convergence thus stems from the exponential convergence of the Hilbert distance $d(e_k, \hat{e}_k)$. ■

IV. Gain tuning issues

As presented in the previous section, the advantage of using correction terms L_k over the method proposed in [3] is that convergence speed can be managed. The lower the projective diameter of $A_k + L_k C_k$ is, the faster the convergence speed is. The goal is to find a $n \times p$ matrix L_k which must satisfy the two linear inequalities (a) - (b) and such that $\Delta(A_k + L_k C_k)$ is the lowest. At least this should be smaller than the value of $\Delta(A)$, the ideal case being for $\Delta(A_k + L_k C_k) = 0$.

A. Optimization problem involved

To simplify exposition, we omit the subscripts of A_k, B_k, L_k and we also assume a scalar output $y_k \in \mathbb{R}_+$, ie. $C \in \mathbb{R}^{1\times n}$. Given A, C at each time step k in order to optimize the contraction rate, one can consider the following problem:

$$\begin{aligned} \underset{L}{\text{minimize}} \quad & \Delta(A+LC) \\ \text{subject to} \quad & (\alpha), (\beta). \end{aligned}$$

where L satisfies

(α) $\ell_{i1} y \leq \sum_{k=1}^{m} b_{ik} u_{k1}$ for all $i \in \{1, ..., n\}$.

(β) $a_{ij} + \ell_{i1} c_{1j} > 0$ for all $i \in \{1, ..., n\}$.

and where

$$\Delta(A+LC) = \max_{i,j,k,r}\left\{\log\left(\frac{(a_{ij}+\ell_{i1}c_{1j})(a_{kr}+l_{k1}c_{1r})}{(a_{ir}+\ell_{i1}c_{1r})(a_{kj}+\ell_{k1}c_{1j})}\right)\right\}. \quad (11)$$

Note that the optimization problem is always feasible with $L = 0$. It is a minmax optimization problem, and can be handled numerically via standard optimization methods, including brute force algorithms if the dimension of the state space is not too large.

B. A geometrical method to decrease $\Delta(A + LC)$

When the matrices A_k are known in advance, typically in the time-invariant case, the optimization problem which consists of minimizing $\Delta(A_k + L_k C_k)$ at each step k can be handled numerically, and the optimal gain matrices L_k can be computed offline. However, when the system is time-varying and the matrix sequence $(A_k)_{k\geq 0}$ is supposed not to be known in advance, for the observer to work in real-time the gain matrices L_k must be computed online.

Computational power limitations can forbid the use of costly optimization methods online. We propose a simple geometric heuristic that guarantees at each step (omitting the subscipt k) $\Delta(A + LC) \leq \Delta(A)$. This implies that with correction terms the error equation is proved to converge faster than without such terms. The method is simple and can be suboptimal, however it proves to be optimal in the two-dimensional case, as shown in the following subsection.

For the sake of illustration, suppose that the scalar output measures the first coordinate of the state vector, i.e., $C_k = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{1\times n}$ for all $k \geq 0$. The idea is then as follows: consider the orthogonal projection of the first column $a_{\bullet,1}$ onto the linear space spanned by the remaining columns $a_{\bullet,j}$ for $2 \leq j \leq n$. In turn, this vector can be projected onto the convex space $\{a_{\bullet,1} + l_{\bullet,1}\}$, where L satisfies conditions $(\alpha), (\beta)$. The corresponding correction matrix L lies in the feasible set, and the angle between the vector $a_{\bullet,1} + l_{\bullet,1}$ and any remaining column $a_{\bullet,j}$ for $2 \leq j \leq n$ is less than the angle between $a_{\bullet,1}$ and $a_{\bullet,j}$. But the function $y \to d(x, y)$ for fixed x is an increasing function of the angle between x and y. As a result, the Hilbert distance between two arbitrary columns of $A + LC$ is less than the distance between the corresponding columns in A. Recalling the characterization of the projective diameter in previous sections, this implies that $\Delta(A + LC) \leq \Delta(A)$.

C. Two-dimensional linear case

In this section, we propose an explicit construction of positive observers for two-dimensional systems. The proposed observer maximizes the rate of contraction as measured by the Hilbert projective metric. It turns out to be a dead-beat observer. Once again we assume $y = x_1$ for simplicity.

Proposition 3: Consider the time varying positive system

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k \\ y_k &= C x_k \end{aligned} \quad (12)$$

where A_k is a matrix in $\mathbb{R}^{2\times 2}$ with positive entries, B_k is a non-negative matrix in $\mathbb{R}^{2\times m}$ and $C = \begin{pmatrix} 1 & 0 \end{pmatrix}$. Then there exists a non-positive 2×1 matrix L_k such that the observer (10) is dead-beat. We have indeed $\hat{x}_k = x_k$ for all $k \geq 1$.

Proof: At each time step k, one can always choose a matrix $L = \begin{pmatrix} \ell_{11} \\ \ell_{21} \end{pmatrix} \in \mathbb{R}^{2\times 1}$ such that

$$-a_{11} < \ell_{11} < \min\left(\frac{-\det(A)}{a_{22}}, 0\right) \quad (13)$$

$$\ell_{21} = \frac{\det(A) + \ell_{11}a_{22}}{a_{12}} \quad (14)$$

where $A = (a_{i,j}) > 0$ for $i, j \in \{1, 2\}$. Now we verify the two conditions $(\alpha) - (\beta)$

- (α) From (13), we have $\ell_{11} < 0$. From the fact that $\ell_{11} < -\frac{\det(A)}{a_{22}}$ and combining with (14), we can deduce that $\ell_{21} < 0$. Consequently $L_k y_k \leq B_k u_k$ at any instant k where u_k is a positive control.
- (β) From (13) and (14), we can easily deduce that $a_{11} + \ell_{11},\ a_{21} + \ell_{21} > 0$. Then $A_k + L_k C_k > 0$ at any time $k \geq 0$.

Furthermore, we have $\det(A + LC) = (a_{11} + \ell_{11})a_{22} - (a_{21} + \ell_{21})a_{12} = 0$. Thus, rank$(A + LC) = 1$. Then $\Delta(A_k + L_k C_k) = 0$ for all $k \in \mathbb{N}$. The optimization problem is solved. As a result, $\|e_k\| = \mu_k$. Then $d(x_k, \hat{x}_k) = 0$ and $\left|\frac{\|x_k\|}{\|\hat{x}_k\|} - 1\right| = 0$ for $k \geq 1$. The proposition follows. ■

V. Numerical example

Consider for example the discrete-time system:

$$x_{k+1} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} x_k \ , \quad y_k = \begin{pmatrix} 1 & 0 \end{pmatrix} x_k \quad (15)$$

We have $\det(A) = 1$. From (13) and (14), $\ell_{11} = -2$, $\ell_{21} = -1$ can be taken. The following conditions are obviously satisfied:

- (α) $A + LC = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is a positive matrix.
- (β) $L = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$ is a negative matrix. Thus, $Ly_k < Bu_k$ for all $k \in \mathbb{N}$.

Moreover $\det(A + LC) = 0$ proving $\Delta(A + LC) = 0$. A dead bit observer is therefore readily obtained:

$$\hat{x}_k = w_k + \frac{x_{1,k} - w_{1,k}}{\hat{e}_{1,k}} \hat{e}_k \quad (16)$$

where

$$\begin{pmatrix} w_{1,k+1} \\ w_{2,k+1} \end{pmatrix} = \begin{pmatrix} w_{1,k} + w_{2,k} + 2x_{1,k} \\ w_{1,k} + w_{2,k} + x_{1,k} \end{pmatrix} \quad (17)$$

and

$$\begin{pmatrix} \hat{e}_{1,k+1} \\ \hat{e}_{2,k+1} \end{pmatrix} = \begin{pmatrix} \hat{e}_{1,k} + \hat{e}_{2,k} \\ \hat{e}_{1,k} + \hat{e}_{2,k} \end{pmatrix}. \quad (18)$$

Figure 1 below illustrates results of observer (16) for system (15). A trajectory with $x_0 = [11/10, 20]^\top$, $w_0 = [1/15, 1]^\top$, $e_0 = [31/30, 19]^\top$ and $\hat{e}_0 = [1, 10]^\top$ as initial conditions is simulated. The top plot 1 shows that the ratio between the norms of x_k and its estimate $\hat{x}_k$ are equal to 1 after a single iteration, whereas the bottom plot shows that their the Hilbert distance between their directions $d(x_k, \hat{x}_k)$ is null after one iteration also, thanks to the convergence of $d(e_k, \hat{e}_k)$.

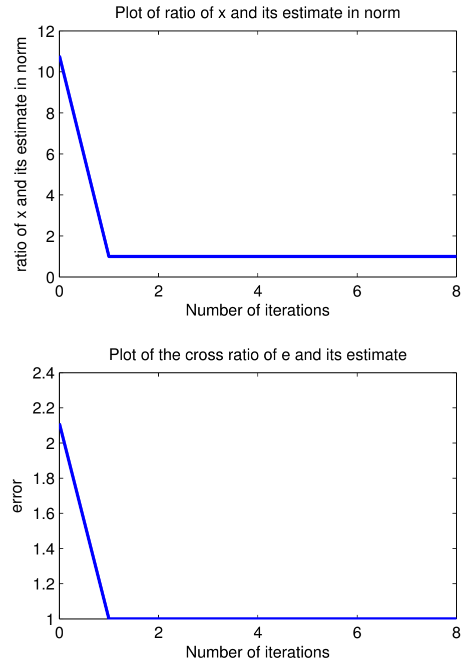

Fig. 1. Top: plot of the quantity $\frac{\|\hat{x}_k\|}{\|x_k\|}$. Bottom: plot of the cross ratio $\frac{\hat{e}_{2k} e_{1k}}{\hat{e}_{1k} e_{2k}}$

Remark 1: An interesting remark concerns the dynamical behavior of the error system: note that, the matrix $A + LC$ is not stable. Thus the corresponding Luenberger observer for (12) is not stable, whereas our technique allows a dead bit estimation of the state. This illustrates well the difference between our approach based on a two-step estimation where first the direction must be estimated, independently from the norm that can grow exponentially. We believe that this approach to estimation problems in the positive orthant indeed suits more the geometry of the state space.

VI. Conclusion

In this paper, we proposed a new design method for positive observers in the positive orthant. The observer design mimicks the Luenberger observer design except that the gain is tuned to maximize the contraction rate of the error as measured by the Hilbert projective metric rather than the usual Euclidean metric that we believe does not suit ideally the geometry of the considered state space. This leads to a counterintuitive result that the observer can converge exponentially although the error system is unstable.

The key ingredient in our approach is the Hilbert metric, that distorts the distances in the orthant so that the boundary is at an infinite distance of any point, leading to a natural preservation of positiveness. We anticipate interesting extensions in the nonlinear case an in more general cones.

REFERENCES

[1] J. Back and A. Astolfi. Design of positive linear observers for positive linear systems via coordinates transformations and positive realizations. *SIAM Journal of Control and Optimization*, 47(1):345–373, 2008.

[2] L. Benvenuti and L. Farina. A tutorial on the positive realization problem. *IEEE Trans. Automat. Control*, 49(5):651–664, 2004.

[3] Silvere Bonnabel, Alessandro Astolfi, and Rodolphe Sepulchre. Contraction and observer design on cones. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 7147–7151. IEEE, 2011.

[4] P.J. Bushell. Hilbert's metric and positive contraction mappings in a banacb space. *Archive for Rational Mechanics and Analysis*, 1973.

[5] L. Farina and S. Rinaldi. *Positive Linear Systems: Theory and Applications*. Wiley, New York, 2000.

[6] H.M Hardïn and J.H van Schuppen. Observers for linear positive systems. *Linear Algebra and its Applications*, 425(2–3):571–607, 2007.

[7] J.A. Jacquez and C.P. Simson. Qualitative theory of compartmental systems. *SIAM Reviews*, 35:43–79, 1993.

[8] W. Lohmiller and J.J.E. Slotine. On contraction analysis for nonlinear systems. *Automatica*, 34(6):683–696, 1998.

[9] Y. Ohta, H. Maeda, and S. Kodama. Reachability, observability and realizability of continuous-time positive systems. *SIAM Journal of Control and Optimization*, 22(2):171–180, 1984.

[10] H.L. Smith. *Monotone Dynamical Systems-An introduction to the theory of competitive and cooperative systems*. American mathematical society, Province, RI, 1995.

[11] J.M. van den Hof. Positive linear observers for linear compartmental systems. *SIAM Journal of Control and Optimization*, 36(2):590–608, 1998.