



HAL
open science

Necessary and sufficient conditions for invariance of convex sets for discrete-time saturated systems

Mirko Fiacchini, Christophe Prieur, Sophie Tarbouriech

► To cite this version:

Mirko Fiacchini, Christophe Prieur, Sophie Tarbouriech. Necessary and sufficient conditions for invariance of convex sets for discrete-time saturated systems. CDC 2013 - 52nd IEEE Conference on Decision and Control, Dec 2013, Florence, Italy. pp.n/c. hal-00920892

HAL Id: hal-00920892

<https://hal.science/hal-00920892>

Submitted on 19 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Necessary and sufficient conditions for invariance of convex sets for discrete-time saturated systems

Mirko Fiacchini¹, Christophe Prieur¹ and Sophie Tarbouriech²

Abstract—A convex analysis-based characterization of invariance and contractivity of compact convex sets for discrete-time saturated systems is presented. Necessary and sufficient conditions for the existence of convex set-induced Lyapunov functions is provided. The results generalize the quadratic Lyapunov theory for saturated systems, apply also to asymmetric saturations and can be extended to affine nonlinearity maps. A numerical example illustrates the improvements of our method with respect to other classical ones.

I. INTRODUCTION

The paper studies discrete-time systems with input saturation. This nonlinearity can degrade the performance of the linear behavior and even destabilize the systems for large initial conditions, see [14], [15]. It is then important to analyze the stability and convergence properties of such systems, for instance through Lyapunov functions induced by invariant and contractive sets. Concerning invariance, see the recent monograph [3]. Different set-induced Lyapunov functions for saturated and nonlinear systems have been considered in literature: quadratic Lyapunov functions and invariant ellipsoids are used in [2], [6], [8], [16]; invariant polytopes are employed in [1], [4], [7].

In this paper, necessary and sufficient conditions are stated for a compact convex set to be invariant or contractive. Such conditions allow also to implicitly characterize local set-induced Lyapunov functions. The paper extends the results in [5], that deals only with symmetric saturations and sets and with sufficient results. The necessary and sufficient condition holds also for asymmetric saturations and it can be extended to systems with affine nonlinearity at the input. The particularization of the results to ellipsoids and quadratic functions permits to compare it with analogous methods, see [2], [9]–[12], and to highlight the substantial improvement achieved. In fact, their conservatism is overcome in our method, as shown through a simple one-dimensional illustrative example. The example provides the evidence of how conservative can be the only sufficient conditions for invariance, also when considering very low dimensional dynamics and very simple and common sets, as ellipsoids.

The paper is organized as follows: Section II presents the problem statement. The main result is stated in Section III where the necessary and sufficient conditions for a compact, convex set to be contractive are given. The comparison with

other methods and a discussion on the obtained results are given in Section IV. Section V collects some conclusions.

Notation: Given $n \in \mathbb{N}$, define $\mathbb{N}_n = \{x \in \mathbb{N} : 1 \leq x \leq n\}$. Given $A \in \mathbb{R}^{n \times m}$, A_i with $i \in \mathbb{N}_n$ denotes its i -th row, $A_{(j)}$ with $j \in \mathbb{N}_m$ its j -th column. Given $\Omega \subseteq \mathbb{R}^n$ and $\alpha \geq 0$, define the set $\alpha\Omega = \{\alpha x \in \mathbb{R}^n : x \in \Omega\}$, the interior of Ω is $\text{int}(\Omega)$ and its boundary is $\partial\Omega$. The left-hand side and right-hand side of an equation or inequality are shorten as l.h.s. and r.h.s.

II. PROBLEM STATEMENT

Consider discrete-time linear saturated systems given by

$$x^+ = f(x) = Ax + B\varphi(Kx), \quad (1)$$

where $x \in \mathbb{R}^n$ is the current state, $x^+ \in \mathbb{R}^n$ is the successor and the saturated feedback control is given by $u = \varphi(Kx) \in \mathbb{R}^m$. Function $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}^m$ denotes the saturation function which is defined by $\varphi_i(y) = \text{sgn}(y_i) \min\{|y_i|, 1\}$ for every $i \in \mathbb{N}_m$.

Remark 1: Although we only consider saturated functions, the results presented are directly extendable to the case of piecewise affine functions, since they can be expressed in form of saturation functions, see for instance [12].

It is useful to introduce the concept of support function.

Definition 1: Given a set $\Omega \subseteq \mathbb{R}^n$, the support function of Ω evaluated at $\eta \in \mathbb{R}^n$ is $\phi_\Omega(\eta) = \sup_{x \in \Omega} \eta^T x$.

Geometrically, the support function of Ω at η is the signed “distance” of the point of the closure of Ω further from the origin, along the direction η . See [13] for some properties of support functions. Set-inclusion conditions can be given in terms of linear inequalities involving the support functions, as recalled here (see [13], for instance).

Property 1: Given two closed, convex sets $\Omega \subseteq \mathbb{R}^n$ and $\Gamma \subseteq \mathbb{R}^n$, then $x \in \Omega$ if and only if $\eta^T x \leq \phi_\Omega(\eta)$ for all $\eta \in \mathbb{R}^n$, and $\Gamma \subseteq \Omega$ if and only if $\phi_\Gamma(\eta) \leq \phi_\Omega(\eta)$, for all $\eta \in \mathbb{R}^n$.

The definitions of invariant and λ -contractive sets follows.

Definition 2 ([3]): A set $\Omega \subseteq \mathbb{R}^n$ is an invariant set for the system $x^+ = f(x)$ if $f(x) \in \Omega$, for all $x \in \Omega$.

Every trajectory starting in an invariant Ω remains in it.

Definition 3 ([3]): A convex compact set $\Omega \subseteq \mathbb{R}^n$ with $0 \in \text{int}(\Omega)$ is a λ -contractive set for the system $x^+ = f(x)$ if $f(x) \in \lambda\Omega$, for all $x \in \Omega$, with $\lambda \in [0, 1]$.

Since λ -contractivity implies invariance, when in the following we will guarantee λ -contractivity, we will implicitly ensure also invariance. The property of λ -contractivity of a compact convex set can be used to induce a local Lyapunov function. We are interested here to conditions on compact convex sets $\Omega \subseteq \mathbb{R}^n$, with $0 \in \text{int}(\Omega)$, whose satisfaction ensures that every set $\alpha\Omega$, with $\alpha \in [0, 1]$, is λ -contractive.

¹ GIPSA-lab, Grenoble Campus, 11 rue des Mathématiques, BP 46, 38402 Saint Martin d’Hères Cedex, France. {mirko.fiacchini, christophe.prieur}@gipsa-lab.fr

² CNRS, LAAS, 7 Avenue du Colonel Roche, F31400 Toulouse, France and Univ de Toulouse, LAAS, F31400 Toulouse, France. sophie.tarbouriech@laas.fr

This would imply that there exists a local Lyapunov function defined on Ω , whose level sets are $\alpha\Omega$ with $\alpha \in [0, 1]$.

Given a convex, compact set $\Omega \subseteq \mathbb{R}^n$ with $0 \in \text{int}(\Omega)$, the Minkowski function of Ω at x is defined as

$$\Psi_\Omega(x) = \min_{\alpha \geq 0} \{ \alpha \in \mathbb{R} : x \in \alpha\Omega \}.$$

Intuitively, the value of $\Psi_\Omega(x)$ is how much the set Ω should be scaled for x to be on its boundary, that is such that $x \in \partial(\Psi_\Omega(x)\Omega)$. Then $x \in \partial\Omega(x)$ where we define

$$\Omega(x) = \Psi_\Omega(x)\Omega. \quad (2)$$

The set $\Omega(x)$ is useful to determine the condition for the sets $\alpha\Omega$ to be λ -contractive for the saturated system (1), for all $\alpha \in [0, 1]$. Such a condition is given by a (possibly uncountable) set of nonconvex constraints, as stated below.

Proposition 1: Given the system (1), the convex compact set Ω with $0 \in \text{int}(\Omega)$ is such that $\alpha\Omega$ is λ -contractive for every $\alpha \in [0, 1]$ if and only if

$$\eta^T f(x) \leq \lambda \phi_{\Omega(x)}(\eta), \quad (3)$$

for all $x \in \Omega$ and every $\eta \in \mathbb{R}^n$.

Proof: Sets $\alpha\Omega$ are λ -contractive for every $\alpha \in [0, 1]$ if and only if $x^+ \in \lambda\Omega(x)$, for all $x \in \Omega$. This is equivalent, by Property 1, to (3) for every $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^n$. ■

The λ -contractivity of sets $\alpha\Omega$ is equivalent to the decreasing of $\Psi_\Omega(x)$, then it implies local convergence in Ω .

III. NECESSARY AND SUFFICIENT CONDITION FOR THE λ -CONTRACTIVITY OF A CONVEX SET

Given the system (1) and $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^n$ we define

$$\begin{aligned} \mathcal{N}^+(x, \eta) &= \{i \in \mathbb{N}_m : \eta^T B_{(i)} > 0, K_i x < -1\}, \\ \mathcal{N}^-(x, \eta) &= \{i \in \mathbb{N}_m : \eta^T B_{(i)} < 0, K_i x > 1\}, \\ \mathcal{N}(x, \eta) &= \mathcal{N}^+(x, \eta) \cup \mathcal{N}^-(x, \eta), \\ \mathcal{P}^+(x, \eta) &= \{i \in \mathbb{N}_m : \eta^T B_{(i)} > 0, K_i x > 1\}, \\ \mathcal{P}^-(x, \eta) &= \{i \in \mathbb{N}_m : \eta^T B_{(i)} < 0, K_i x < -1\}, \\ \mathcal{P}(x, \eta) &= \mathcal{P}^+(x, \eta) \cup \mathcal{P}^-(x, \eta), \\ \mathcal{L}^+(x, \eta) &= \{i \in \mathbb{N}_m : \eta^T B_{(i)} > 0, |K_i x| \leq 1\}, \\ \mathcal{L}^-(x, \eta) &= \{i \in \mathbb{N}_m : \eta^T B_{(i)} < 0, |K_i x| \leq 1\}, \\ \mathcal{L}(x, \eta) &= \mathcal{L}^+(x, \eta) \cup \mathcal{L}^-(x, \eta), \\ \mathcal{O}(\eta) &= \{i \in \mathbb{N}_m : \eta^T B_{(i)} = 0\}. \end{aligned} \quad (4)$$

The sets defined in (4), subsets of \mathbb{N}_m , permit to characterize the regions of $(x, \eta) \in \mathbb{R}^n \times \mathbb{R}^n$ where the input saturates and the different values of $\eta^T B_{(i)} \phi_i(Kx)$ in these regions. In fact:

$$\begin{cases} \eta^T B_{(i)} \phi_i(Kx) = -\eta^T B_{(i)}, & \text{if } i \in \mathcal{N}^+(x, \eta), \\ \eta^T B_{(i)} \phi_i(Kx) = \eta^T B_{(i)}, & \text{if } i \in \mathcal{N}^-(x, \eta), \\ \eta^T B_{(i)} \phi_i(Kx) = \eta^T B_{(i)}, & \text{if } i \in \mathcal{P}^+(x, \eta), \\ \eta^T B_{(i)} \phi_i(Kx) = -\eta^T B_{(i)}, & \text{if } i \in \mathcal{P}^-(x, \eta), \\ \eta^T B_{(i)} \phi_i(Kx) = \eta^T B_{(i)} K_i x, & \text{if } i \in \mathcal{L}(x, \eta), \\ \eta^T B_{(i)} \phi_i(Kx) = 0, & \text{if } i \in \mathcal{O}(\eta), \end{cases} \quad (5)$$

for every $i \in \mathbb{N}_m$. Define an equivalence relation \sim_x on \mathbb{R} denoting $y \sim_x z$ if and only if either: $y > 1$ and $z > 1$; or $y < -1$ and $z < -1$; or $|y| \leq 1$ and $|z| \leq 1$. A second equivalence relation \sim_η on \mathbb{R} is defined saying that $d, e \in \mathbb{R}$ are such that $d \sim_\eta e$ if and only if either: $d > 0$ and $e > 0$; or $d < 0$

and $e < 0$; or $d = 0$ and $e = 0$. We define two partitions of \mathbb{R}^n induced by the equivalence relations \sim_x and \sim_η as

$$\begin{aligned} \mathcal{J} &= \{J \subseteq \mathbb{R}^n : x, \bar{x} \in J \Leftrightarrow x_i \sim_x \bar{x}_i, \forall i \in \mathbb{N}_n\}, \\ \mathcal{K} &= \{E \subseteq \mathbb{R}^n : \eta, \bar{\eta} \in E \Leftrightarrow \eta_i \sim_\eta \bar{\eta}_i, \forall i \in \mathbb{N}_m\}. \end{aligned} \quad (6)$$

Given $J \in \mathcal{J}$ and $E \in \mathcal{K}$, for every $x, \bar{x} \in J$ and $\eta, \bar{\eta} \in E$, the sets in (4) are the same. That is $\mathcal{N}^+(x, \eta) = \mathcal{N}^+(x, \bar{\eta}) = \mathcal{N}^+(\bar{x}, \bar{\eta}) = \mathcal{N}^+(\bar{x}, \eta)$ (analogously, for any other set in (4)). Also relations (5) are the same within $J \times E$, for every $J \in \mathcal{J}$ and $E \in \mathcal{K}$. For every $J \in \mathcal{J}$ and $E \in \mathcal{K}$ and given $x \in J$ and $\eta \in E$, denote, with a slight abuse of notation,

$$\begin{aligned} \mathcal{N}^+(J, E) &= \mathcal{N}^+(x, \eta), & \mathcal{N}^-(J, E) &= \mathcal{N}^-(x, \eta), \\ \mathcal{P}^+(J, E) &= \mathcal{P}^+(x, \eta), & \mathcal{P}^-(J, E) &= \mathcal{P}^-(x, \eta), \\ \mathcal{L}^+(J, E) &= \mathcal{L}^+(x, \eta), & \mathcal{L}^-(J, E) &= \mathcal{L}^-(x, \eta), \\ \mathcal{O}(E) &= \mathcal{O}(\eta). \end{aligned} \quad (7)$$

First a trivial necessary condition is presented.

Proposition 2: Given the system (1), and the compact convex set $\Omega \subseteq \mathbb{R}^n$, with $0 \in \text{int}(\Omega)$, if condition (3) holds for every $x \in \Omega$ and every $\eta \in \mathbb{R}^n$ then $\eta^T (A + BK)x \leq \lambda \phi_{\Omega(x)}(\eta)$, is satisfied for every $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^n$.

Now we provide a necessary and sufficient condition for (3) to hold in Ω , which is the main contribution of the paper.

Theorem 1: Given the system (1), the compact convex set $\Omega \subseteq \mathbb{R}^n$, with $0 \in \text{int}(\Omega)$, is such that $\alpha\Omega$ is λ -contractive for every $\alpha \in [0, 1]$ if and only if for every $J \in \mathcal{J}$ and $E \in \mathcal{K}$, there exist $\gamma_i^{J,E} \in \mathbb{R}$ and $\sigma_i^{J,E}(x) \in \mathbb{R}$, with $i \in \mathbb{N}_m$, such that

$$\eta^T Ax + \sum_{i \in \mathbb{N}_m} \gamma_i^{J,E} \eta^T B_{(i)} K_i x + \sigma_i^{J,E}(x) \eta^T B_{(i)} \leq \lambda \phi_{\Omega(x)}(\eta), \quad (8)$$

and

$$\begin{cases} \gamma_i^{J,E} \leq 0, & \sigma_i^{J,E}(x) \geq \gamma_i^{J,E} - 1, & \text{if } i \in \mathcal{N}^+(J, E), \\ \gamma_i^{J,E} \leq 0, & \sigma_i^{J,E}(x) \leq 1 - \gamma_i^{J,E}, & \text{if } i \in \mathcal{N}^-(J, E), \\ \gamma_i^{J,E} \geq 0, & \sigma_i^{J,E}(x) \geq 1 - \gamma_i^{J,E}, & \text{if } i \in \mathcal{P}^+(J, E), \\ \gamma_i^{J,E} \geq 0, & \sigma_i^{J,E}(x) \leq \gamma_i^{J,E} - 1, & \text{if } i \in \mathcal{P}^-(J, E), \\ \gamma_i^{J,E} \in \mathbb{R}, & \begin{cases} \sigma_i^{J,E}(x) \geq \gamma_i^{J,E} - 1, \\ \sigma_i^{J,E}(x) \geq 1 - \gamma_i^{J,E}, \end{cases} & \text{if } i \in \mathcal{L}^+(J, E), \\ \gamma_i^{J,E} \in \mathbb{R}, & \begin{cases} \sigma_i^{J,E}(x) \leq \gamma_i^{J,E} - 1, \\ \sigma_i^{J,E}(x) \leq 1 - \gamma_i^{J,E}, \end{cases} & \text{if } i \in \mathcal{L}^-(J, E), \\ \gamma_i^{J,E} \in \mathbb{R}, & \sigma_i^{J,E}(x) \in \mathbb{R}, & \text{if } i \in \mathcal{O}(E), \end{cases} \quad (9)$$

hold for all $x \in \Omega$ and all $\eta \in \mathbb{R}^n$.

Proof: The proof is divided in two main parts, the first concerning the sufficiency and the second one the necessity.

Sufficiency: Given the sets $J \in \mathcal{J}$ and $E \in \mathcal{K}$, consider $\eta \in E$ and $x \in J \cap \Omega$. Every index $i \in \mathbb{N}_m$ belongs to one of the sets defined in (4), all the possibilities are considered.

• ($\mathcal{N}^+(J, E)$). Suppose that $i \in \mathcal{N}^+(J, E)$, which implies $\eta^T B_{(i)} \phi_i(Kx) = -\eta^T B_{(i)}$, see (5) and (7). From (4), $\eta \in E$ implies the satisfaction of constraint $\eta^T B_{(i)} > 0$. Then, applying the S-procedure, we have that (3) holds at $x \in J \cap \Omega$ for every $\eta \in E$ if there exists $\tau_i(x) \geq 0$ such that

$$\eta^T Ax + \sum_{j \neq i} \eta^T B_{(j)} \phi_j(Kx) - \eta^T B_{(i)} + \tau_i(x) \eta^T B_{(i)} \leq \lambda \phi_{\Omega(x)}(\eta), \quad (10)$$

for every $\eta \in \mathbb{R}^n$. Applying once again the S-procedure, condition (10) can be imposed for every $x \in J \cap \Omega$. Condition $x \in J$ is equivalent to $\eta^T B_{(i)} K_i x < -\eta^T B_{(i)}$, being $\eta^T B_{(i)} > 0$ for all $\eta \in E$. Then (3) is satisfied for all $x \in J \cap \Omega$ and for every $\eta \in E$ if there exist $\tau_i(x) \geq 0$ and $\delta_i \geq 0$ such that

$$\begin{aligned} & \eta^T A x + \sum_{j \neq i} \eta^T B_{(j)} \varphi_j(Kx) - \eta^T B_{(i)} + \tau_i(x) \eta^T B_{(i)} \\ & + \delta_i (-\eta^T B_{(i)} K_i x - \eta^T B_{(i)}) \leq \lambda \phi_{\Omega(x)}(\eta), \end{aligned} \quad (11)$$

holds for every $x \in \Omega$ and $\eta \in \mathbb{R}^n$. Notice that, whereas $\tau_i(x)$ is a function of the state since it concerns a condition imposed for a particular value of x , δ_i is constant since it is related to a condition involving the whole subset J . Finally, we have that (11) with $\tau_i(x) \geq 0$ and $\delta_i \geq 0$ is equivalent to

$$\begin{aligned} & \eta^T A x + \sum_{j \neq i} \eta^T B_{(j)} \varphi_j(Kx) \\ & + \gamma_i^{J,E} \eta^T B_{(i)} K_i x + \sigma_i^{J,E}(x) \eta^T B_{(i)} \leq \lambda \phi_{\Omega(x)}(\eta), \end{aligned} \quad (12)$$

for every $x \in \Omega$ and every $\eta \in \mathbb{R}^n$, with

$$\gamma_i^{J,E} = -\delta_i \leq 0, \quad \sigma_i^{J,E}(x) = -1 + \gamma_i^{J,E} + \tau_i(x) \geq -1 + \gamma_i^{J,E}.$$

• ($\mathcal{N}^-(J, E)$). Similar considerations, for $i \in \mathcal{N}^-(J, E)$, lead to the fact that condition (3) holds at $x \in J \cap \Omega$ for every $\eta \in E$ if there exist $\gamma_i^{J,E}, \sigma_i^{J,E}(x) \in \mathbb{R}$ such that (12) is satisfied for all $x \in \Omega$ and $\eta \in \mathbb{R}^n$ with

$$\gamma_i^{J,E} = -\delta_i \leq 0, \quad \sigma_i^{J,E}(x) \leq 1 - \gamma_i^{J,E}.$$

• ($\mathcal{P}^+(J, E)$) and ($\mathcal{P}^-(J, E)$). Analogously, satisfaction of (12) for every $\eta \in \mathbb{R}^n$ with

$$\gamma_i^{J,E} = \delta_i(x) \geq 0, \quad \sigma_i^{J,E}(x) \geq 1 - \gamma_i^{J,E},$$

if $i \in \mathcal{P}^+(J, E)$, and with

$$\gamma_i^{J,E} = \delta_i(x) \geq 0, \quad \sigma_i^{J,E}(x) \leq -1 + \gamma_i^{J,E},$$

if $i \in \mathcal{P}^-(J, E)$, implies (3) for every $\eta \in E$.

• ($\mathcal{L}^+(J, E)$) and ($\mathcal{L}^-(J, E)$). Concerning $i \in \mathcal{L}^+(J, E)$, the S-procedure is applied to three constraints $\eta^T B_{(i)} > 0$, $\eta^T B_{(i)} K_i x \geq -\eta^T B_{(i)}$ and $\eta^T B_{(i)} K_i x \leq \eta^T B_{(i)}$, then involving three nonnegative multipliers $\tau_i(x)$, $\underline{\delta}_i$ and $\bar{\delta}_i$. Conditions on $\gamma_i^{J,E}$ and $\sigma_i^{J,E}(x)$ for $i \in \mathcal{L}^+(J, E)$ are

$$\begin{cases} \gamma_i^{J,E} = 1 - \bar{\delta}_i + \underline{\delta}_i \in \mathbb{R}, \\ \sigma_i^{J,E}(x) \geq \gamma_i^{J,E} - 1, \quad \sigma_i^{J,E}(x) \geq 1 - \gamma_i^{J,E}, \end{cases}$$

where $\sigma_i^{J,E}(x) = \bar{\delta}_i + \underline{\delta}_i + \tau_i(x)$. Analogously, if $i \in \mathcal{L}^-(J, E)$ conditions on $\gamma_i^{J,E} \in \mathbb{R}$, $\sigma_i^{J,E}(x) \geq 0$, are

$$\gamma_i^{J,E} \in \mathbb{R}, \quad \sigma_i^{J,E}(x) \leq \gamma_i^{J,E} - 1, \quad \sigma_i^{J,E}(x) \leq 1 - \gamma_i^{J,E}.$$

• ($\mathcal{O}^+(J, E)$). Finally, if $i \in \mathcal{O}(E)$ we have $\eta^T B_{(i)} = 0$ and then the term related to the i -th input in (3) is zero. Thus (3) is equivalent to (12) for every $\gamma_i^{J,E} \in \mathbb{R}$ and $\sigma_i^{J,E}(x) \in \mathbb{R}$.

Similar arguments, applied to any $i \in \mathbb{N}_m$, yield to prove that satisfaction of (8) for every $x \in \Omega$ and every $\eta \in \mathbb{R}^n$ with (9) is a sufficient condition for (3) to hold for all $\eta \in E$ and $x \in J \cap \Omega$. It has to be stressed that, given $J \in \mathcal{S}$ and $E \in \mathcal{K}$, conditions (8) with (9) are imposed over the whole space of η and for all $x \in \Omega$, and guarantee the satisfaction of condition (3) on the subsets $J \cap \Omega \subseteq \mathbb{R}^n$ and $E \subseteq \mathbb{R}^n$.

Necessity: We have to prove that the satisfaction of condition (3) for every $x \in \Omega$ and $\eta \in \mathbb{R}^n$ implies the existence of $\gamma_i^{J,E}, \sigma_i^{J,E}(x)$, with $i \in \mathbb{N}_m$, such that (8) and (9) hold for every $x \in \Omega$ and $\eta \in \mathbb{R}^n$, for all $J \in \mathcal{S}$ and $E \in \mathcal{K}$. It is important to remember that there is a pair of constraints (8) and (9) for each $J \in \mathcal{S}$ and $E \in \mathcal{K}$; all of them must hold. As the sets $J \in \mathcal{S}$ and $E \in \mathcal{K}$ form a partition of $\mathbb{R}^n \times \mathbb{R}^n$, condition (3) for all $x \in \Omega$ and $\eta \in \mathbb{R}^n$ is equivalent to

$$\eta^T f(x) \leq \lambda \phi_{\Omega(x)}(\eta), \quad \forall x \in J \cap \Omega, \forall \eta \in E, \quad (13)$$

for every $J \in \mathcal{S}$ and $E \in \mathcal{K}$. We fix generic $J \in \mathcal{S}$ and $E \in \mathcal{K}$ and we prove that the related (8) and (9) hold for all $\hat{x} \in \hat{J} \cap \Omega$ and all $\hat{\eta} \in \hat{E}$, for every $\hat{J} \in \mathcal{S}$ and $\hat{E} \in \mathcal{K}$, with appropriate $\gamma_i^{J,E}$ and $\sigma_i^{J,E}(x)$, with $i \in \mathbb{N}_m$. Below, we consider \hat{x} contained in Ω and analyze the possible cases.

The first possibility is that $\hat{J} = J$ and $\hat{E} = E$, that means that $\hat{x} \in J \cap \Omega$ and $\hat{\eta} \in E$. For sake of simplicity, we first consider $m = 1$ and analyze the cases related only to $\hat{\eta}^T B_{(i)} > 0$, i.e. $i \in \mathcal{N}^+(J, E)$, $i \in \mathcal{P}^+(J, E)$ or $i \in \mathcal{L}^+(J, E)$ with $i = 1$. The other cases are analogous.

• ($\mathcal{N}^+(J, E)$). Suppose that $i \in \mathcal{N}^+(J, E)$ with $\hat{J} = J$ and $\hat{E} = E$. From (4) and (5) we have

$$\hat{\eta}^T B_{(i)} \varphi_i(K\hat{x}) = -\hat{\eta}^T B_{(i)}, \quad \hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)} < 0.$$

From (9) there must exist $\alpha_i^{J,E}(\hat{x}) \geq 0$ such that $\sigma_i^{J,E}(\hat{x}) = -1 + \gamma_i^{J,E} + \alpha_i^{J,E}(\hat{x})$ with $\gamma_i^{J,E} \leq 0$ satisfying (8). Posing $\alpha_i^{J,E}(\hat{x}) = 0$, the l.h.s. of (8) is

$$\begin{aligned} & \hat{\eta}^T A \hat{x} + \gamma_i^{J,E} \hat{\eta}^T B_{(i)} K_i \hat{x} + \sigma_i^{J,E}(\hat{x}) \hat{\eta}^T B_{(i)} \\ & = \hat{\eta}^T A \hat{x} + \hat{\eta}^T B_{(i)} \varphi_i(K\hat{x}) + \gamma_i^{J,E} (\hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)}), \end{aligned}$$

and thus (8) is equivalent to

$$\gamma_i^{J,E} (\hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)}) \leq -\hat{\eta}^T f(\hat{x}) + \lambda \phi_{\Omega(\hat{x})}(\hat{\eta}), \quad (14)$$

whose r.h.s. is nonnegative due to (13). Notice that also the l.h.s. of (14) is given by a nonnegative term, but there always exists $\gamma_i^{J,E} \leq 0$ such that its holds, and hence (8) is satisfied for all $\hat{x} \in J \cap \Omega$ and $\hat{\eta} \in E$.

• ($\mathcal{P}^+(J, E)$). If $i \in \mathcal{P}^+(J, E)$ with $\hat{J} = J$ and $\hat{E} = E$, then

$$\hat{\eta}^T B_{(i)} \varphi_i(K\hat{x}) = \hat{\eta}^T B_{(i)}, \quad \hat{\eta}^T B_{(i)} K_i \hat{x} - \hat{\eta}^T B_{(i)} > 0,$$

and (9) is equivalent to $\sigma_i^{J,E}(\hat{x}) = 1 - \gamma_i^{J,E} + \alpha_i^{J,E}(\hat{x})$ with $\alpha_i^{J,E}(\hat{x}) \geq 0$ and $\gamma_i^{J,E} \geq 0$. For $\alpha_i^{J,E}(\hat{x}) = 0$ the l.h.s. of (8) is

$$\begin{aligned} & \hat{\eta}^T A \hat{x} + \gamma_i^{J,E} \hat{\eta}^T B_{(i)} K_i \hat{x} + \sigma_i^{J,E}(\hat{x}) \hat{\eta}^T B_{(i)} \\ & = \hat{\eta}^T A \hat{x} + \hat{\eta}^T B_{(i)} \varphi_i(K\hat{x}) + \gamma_i^{J,E} (\hat{\eta}^T B_{(i)} K_i \hat{x} - \hat{\eta}^T B_{(i)}), \end{aligned}$$

and (8) results in

$$\gamma_i^{J,E} (\hat{\eta}^T B_{(i)} K_i \hat{x} - \hat{\eta}^T B_{(i)}) \leq -\hat{\eta}^T f(\hat{x}) + \lambda \phi_{\Omega(\hat{x})}(\hat{\eta}), \quad (15)$$

whose l.h.s. and r.h.s. are nonnegative. The inequality (15), then also (8), can be satisfied.

• ($\mathcal{L}^+(J, E)$). If $i \in \mathcal{L}^+(J, E)$ with $\hat{J} = J$ and $\hat{E} = E$ then

$$\hat{\eta}^T B_{(i)} \varphi_i(K\hat{x}) = \hat{\eta}^T B_{(i)} K_i \hat{x}, \quad |\hat{\eta}^T B_{(i)} K_i \hat{x}| \leq \hat{\eta}^T B_{(i)}, \quad \hat{\eta}^T B_{(i)} > 0.$$

Condition (9) holds if and only if there are $\alpha_i^{J,E}(\hat{x}) \geq 0$ and $\gamma_i^{J,E} \in \mathbb{R}$ such that $\sigma_i^{J,E}(\hat{x}) = \max\{1 - \gamma_i^{J,E}, \gamma_i^{J,E} - 1\} + \alpha_i^{J,E}(\hat{x}) = |1 - \gamma_i^{J,E}| + \alpha_i^{J,E}(\hat{x})$. Fixing $\alpha_i^{J,E}(\hat{x}) = 0$ the l.h.s. of (8) is

$$\hat{\eta}^T A \hat{x} + \gamma_i^{J,E} \hat{\eta}^T B_{(i)} K_i \hat{x} + \sigma_i^{J,E}(\hat{x}) \hat{\eta}^T B_{(i)} = \hat{\eta}^T A \hat{x} + \hat{\eta}^T B_{(i)} \varphi_i(K \hat{x}) + (\gamma_i^{J,E} - 1) \hat{\eta}^T B_{(i)} K_i \hat{x} + |1 - \gamma_i^{J,E}| \hat{\eta}^T B_{(i)},$$

and (8) is

$$(\gamma_i^{J,E} - 1) \hat{\eta}^T B_{(i)} K_i \hat{x} + |1 - \gamma_i^{J,E}| \hat{\eta}^T B_{(i)} \leq -\hat{\eta}^T f(\hat{x}) + \lambda \phi_{\Omega(\hat{x})}(\hat{\eta}). \quad (16)$$

Also in this case the l.h.s. of (16) is nonnegative since

$$-(\gamma_i^{J,E} - 1) \hat{\eta}^T B_{(i)} K_i \hat{x} \leq |\gamma_i^{J,E} - 1| |\hat{\eta}^T B_{(i)} K_i \hat{x}| \leq |1 - \gamma_i^{J,E}| |\hat{\eta}^T B_{(i)}|,$$

and the r.h.s. as well from (13). Nevertheless, there is a $\gamma_i^{J,E} \in \mathbb{R}$, close enough to 1, such that (16) holds and then also (8).

• ($\mathcal{O}(J, E)$). This case can be proved by continuity of the function $\phi_{\Omega(x)}(\eta)$, with respect to both x and η . Alternatively, it is sufficient to notice that (3) and (8) are equivalent if $\eta^T B_{(i)} = 0$, for every $\gamma_i^{J,E}$ and σ_i .

Summarizing: if $m = 1$ and (13) holds for all J and E , then also (8) and (9) can be satisfied for every J and E , provided $\hat{x} \in J \cap \Omega$ and $\hat{\eta} \in E$.

What is left to be proved is that such an implication is satisfied also for $\hat{x} \notin J$ and/or $\hat{\eta} \notin E$ (but $\hat{x} \in \Omega$). Then suppose that (13) holds for all pair $J \in \mathcal{J}$ and $E \in \mathcal{E}$ and consider $\hat{x} \in \hat{J} \cap \Omega$ and $\hat{\eta} \in \hat{E}$ with $\hat{J} \neq J$ and/or $\hat{E} \neq E$. Also in this case we consider at first $m = 1$. We consider J and E related to the region determined by $\eta^T B_{(i)} > 0$ and $K_i x < -1$, i.e. the only nonempty set is $\mathcal{N}^+(J, E)$. As proved above, there exists $\gamma_i^{J,E} \leq 0$ such that (14), and thus also (8), is satisfied for all $x \in J \cap \Omega$ and $\eta \in E$. We denote with $\hat{\gamma}_i^{J,E}$ such an admissible nonpositive value.

It is important to remember that, as (8) and (9) must hold for the given J and E , then the constraint on $\sigma_i^{J,E}$ is $\sigma_i^{J,E}(x) \geq -1 + \hat{\gamma}_i^{J,E}$, which is equivalent to the existence of $\alpha_i^{J,E}(x) \geq 0$ such that $\sigma_i^{J,E}(x) = -1 + \hat{\gamma}_i^{J,E} + \alpha_i^{J,E}(x)$.

• Cases $i \in \mathcal{N}^-(\hat{x}, \hat{\eta})$, $i \in \mathcal{D}^-(\hat{x}, \hat{\eta})$ or $i \in \mathcal{L}^-(\hat{x}, \hat{\eta})$. Assuming that $i \in \mathcal{N}^-(\hat{x}, \hat{\eta})$, $i \in \mathcal{D}^-(\hat{x}, \hat{\eta})$ or $i \in \mathcal{L}^-(\hat{x}, \hat{\eta})$, implies $\hat{\eta}^T B_{(i)} < 0$ in every case. Then, given $\hat{\gamma}_i^{J,E} \leq 0$, for every $\sigma_i^{J,E}(\hat{x}) = -1 + \hat{\gamma}_i^{J,E} + \alpha_i^{J,E}(\hat{x})$ we have that (8) is equivalent to

$$\hat{\eta}^T A \hat{x} - \hat{\eta}^T B_{(i)} + \gamma_i^{J,E} (\hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)}) + \alpha_i^{J,E}(\hat{x}) \hat{\eta}^T B_{(i)} \leq \lambda \phi_{\Omega(\hat{x})}(\hat{\eta}), \quad (17)$$

which is satisfied by an appropriate choice of $\alpha_i^{J,E}(\hat{x})$, since $\hat{\eta}^T B_{(i)} < 0$ and $\alpha_i^{J,E}(\hat{x}) \geq 0$. Notice in fact that the l.h.s. of (17) can be made arbitrarily small.

• Case $i \in \mathcal{D}^+(\hat{x}, \hat{\eta})$. Assume that $i \in \mathcal{D}^+(\hat{x}, \hat{\eta})$, which means that $\hat{\eta}^T B_{(i)} > 0$ and $K_i \hat{x} > 1$, and thus:

$$\begin{cases} \hat{\eta}^T B_{(i)} \varphi_i(K \hat{x}) = \hat{\eta}^T B_{(i)}, \\ \hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)} > \hat{\eta}^T B_{(i)} K_i \hat{x} - \hat{\eta}^T B_{(i)} > 0. \end{cases}$$

Posing $\sigma_i^{J,E}(\hat{x}) = -1 + \hat{\gamma}_i^{J,E} + \alpha_i^{J,E}(\hat{x})$ and $\alpha_i^{J,E}(\hat{x}) = 2$, condition (8) results in

$$\hat{\eta}^T A \hat{x} + \hat{\eta}^T B_{(i)} + \hat{\gamma}_i^{J,E} (\hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)}) \leq \lambda \phi_{\Omega(\hat{x})}(\hat{\eta}), \quad (18)$$

that is equivalent to (14), which is satisfied since $\hat{\gamma}_i^{J,E} (\hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)}) \leq 0$ and $-\hat{\eta}^T f(\hat{x}) + \lambda \phi_{\Omega(\hat{x})}(\hat{\eta}) \geq 0$, from condition (13).

• Case $i \in \mathcal{L}^+(\hat{x}, \hat{\eta})$. In this case we have that $\hat{\eta}^T B_{(i)} > 0$, $|K_i \hat{x}| \leq 1$ and then

$$\hat{\eta}^T B_{(i)} \varphi_i(K \hat{x}) = \hat{\eta}^T B_{(i)} K_i \hat{x}, \quad \hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)} \geq 0.$$

Defining $\sigma_i^{J,E}(\hat{x}) = -1 + \hat{\gamma}_i^{J,E} + \alpha_i^{J,E}(\hat{x})$ and $\alpha_i^{J,E}(\hat{x}) = K_i \hat{x} + 1$, which implies $\alpha_i^{J,E}(\hat{x}) \geq 0$, condition (8) reads

$$\hat{\eta}^T A \hat{x} + \hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\gamma}_i^{J,E} (\hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)}) \leq \lambda \phi_{\Omega(\hat{x})}(\hat{\eta}), \quad (19)$$

which is equivalent to (14). From the signs of $\hat{\gamma}_i^{J,E}$ and $\hat{\eta}^T B_{(i)} K_i \hat{x} + \hat{\eta}^T B_{(i)}$ and condition (13), then (8) is satisfied by an admissible $\sigma_i^{J,E}(\hat{x})$.

• Case $i \in \mathcal{O}(\hat{\eta})$. The proof comes from the equivalence of (3) and (8) for $\hat{\eta}^T B_{(i)} = 0$, for every $\hat{\gamma}_i^{J,E}$ and $\sigma_i^{J,E}(x)$.

Let us summarize the result for the case $m = 1$. Considering $J \in \mathcal{J}$ and $E \in \mathcal{E}$ we discriminated three cases:

1) if $\hat{x} \in J$ and $\hat{\eta} \in E$, admissible $\hat{\gamma}_i^{J,E}$ and $\sigma_i^{J,E}(\hat{x})$ can be determined such that (8) and (9) hold, using condition (13);

2) if $\hat{\eta} \notin E$, then an admissible choice of $\sigma_i^{J,E}(x)$ can be done such that the l.h.s. of (8) is arbitrarily small, for every $\hat{\gamma}_i^{J,E}$. Then, (8) holds;

3) if $\hat{\eta} \in E$ but $\hat{x} \notin J$, an appropriate choice of $\sigma_i^{J,E}(\hat{x})$ leads the l.h.s. of (8) to be equal to $\hat{\eta}^T f(\hat{x})$ plus a nonpositive term. Thus, using condition (13), satisfaction of (8) is proved.

The proof for the case of $m > 1$ is based on the same reasoning as for $m = 1$. We avoid the tedious details, providing just a sketch of the proof. Suppose that $m > 1$ and that (13) holds for all $J \in \mathcal{J}$ and $E \in \mathcal{E}$. Consider a pair $J \in \mathcal{J}$ and $E \in \mathcal{E}$ and generic $\hat{x} \in \Omega$ and $\hat{\eta} \in \mathbb{R}^n$. Three substantially different cases are possible.

(i) $\hat{x} \in J \cap \Omega$ and $\hat{\eta} \in E$. There exist $\gamma_i^{J,E}$ and $\sigma_i^{J,E}(\hat{x})$ satisfying (9), with $i \in \mathbb{N}_m$, such that (8) results in

$$\hat{\eta}^T f(\hat{x}) + \sum_{i \in \mathbb{N}_m} h_i(\gamma_i^{J,E}, \hat{x}, \hat{\eta}) \leq \lambda \phi_{\Omega(\hat{x})}(\hat{\eta}), \quad (20)$$

with functions $h_i(\gamma_i^{J,E}, \hat{x}, \hat{\eta})$ nonnegative, continuous in $\gamma_i^{J,E}$ and vanishing for some admissible value of $\gamma_i^{J,E}$ (i.e. the l.h.s. functions in (14), (15) and (16) are vanishing. The first two l.h.s. functions vanish at the origin whereas the third vanishes at 1). Then, from (13), there are admissible $\gamma_i^{J,E}$, with $i \in \mathbb{N}_m$, such that $\sum_{i \in \mathbb{N}_m} h_i(\gamma_i^{J,E}, \hat{x}, \hat{\eta})$ is small enough for (20) to be satisfied.

(ii) $\hat{\eta} \notin E$. This means that there exists $j \in \mathbb{N}_m$ such that the signs of $\hat{\eta}^T B_{(j)}$ and of $\eta^T B_{(j)}$, for $\eta \in E$, are different (we can neglect the case of $\hat{\eta}^T B_{(j)} = 0$). In this case there is an admissible $\sigma_j^{J,E}(\hat{x})$ such that the j -th term in l.h.s. of (8) is arbitrarily small (e.g. inequality (17)) and then (8) can be satisfied.

(iii) $\hat{\eta} \in E$ but $\hat{x} \notin J$. In this case, two sets of indices can be defined: $I = \{i \in \mathbb{N}_m : K_i \hat{x} \sim_x K_i x, \forall x \in J\}$ and $\bar{I} = J/I$, and there are admissible $\sigma_i^{J,E}(\hat{x})$ such that (8) becomes

$$\hat{\eta}^T f(\hat{x}) + \sum_{i \in I} h_i(\gamma_i^{J,E}, \hat{x}, \hat{\eta}) + \sum_{i \in \bar{I}} g_i(\gamma_i^{J,E}, \hat{x}, \hat{\eta}) \leq \lambda \phi_{\Omega(\hat{x})}(\hat{\eta}), \quad (21)$$

where $h_i(\gamma_i^{J,E}, \hat{x}, \hat{\eta})$ are as defined in (i) and $g_i(\gamma_i^{J,E}, \hat{x}, \hat{\eta})$ are nonpositive functions (see for instance (18) and (19)). From the sign of $g_i(\gamma_i^{J,E}, \hat{x}, \hat{\eta})$, condition (21) holds if $\hat{\eta}^T f(\hat{x}) + \sum_{i \in I} h_i(\gamma_i^{J,E}, \hat{x}, \hat{\eta}) \leq \lambda \phi_{\Omega(\hat{x})}(\hat{\eta})$, which can be satisfied, as for (20), by choosing appropriate $\gamma_i^{J,E}$ and from (13).

Thus finally, if (3) (or equivalently (13) for all $J \in \mathcal{J}$ and $E \in \mathcal{K}$) holds for every $x \in \Omega$ and every $\eta \in \mathbb{R}^n$, then admissible $\gamma_i^{J,E}$ and $\sigma_i^{J,E}$ with $i \in \mathbb{N}_m$ exist such that the constraints of the type (20) and (21) can be satisfied. This implies that (8) and (9) admit a solution for all $J \in \mathcal{J}$ and $E \in \mathcal{K}$ also for $m > 1$, provided that (3) holds for $x \in \Omega$. ■

Theorem 1 establishes a necessary and sufficient condition for the existence of a local exponential set-induced Lyapunov functions for saturated systems.

Remark 2: Besides the theoretical contribution, we think that Theorem 1 can lead to computational tools useful for the analysis of stability of saturated systems. For instance, considering ellipsoidal and polytopic sets Ω , the conditions (8) and (9) could results in convex constraints, computationally tractable. Moreover, notice that the knowledge about which are the nonempty sets $J \in \mathcal{J}$ and $E \in \mathcal{K}$ seems to be required. Nevertheless, we think that the selection of nonempty sets $J \cap \Omega$ and E is implicitly assured by using Theorem 1 (see the numerical example). These appealing properties are the objective of our future research.

A more computation-oriented sufficient condition for the hypotheses of Theorem 1 to hold, concerning the existence of affine functions $\sigma_i^{J,E}(x)$, follows.

Proposition 3: Given the system (1), the compact convex set $\Omega \subseteq \mathbb{R}^n$, with $0 \in \text{int}(\Omega)$, is such that $\alpha\Omega$ is λ -contractive for every $\alpha \in [0, 1]$ if for every $J \in \mathcal{J}$ and every $E \in \mathcal{K}$, there exist $\gamma_i^{J,E} \in \mathbb{R}$, $G_i^{J,E} \in \mathbb{R}^{1 \times n}$ and $\mu_i^{J,E} \in \mathbb{R}$, with $i \in \mathbb{N}_m$, such that (8) and (9) hold with $\sigma_i^{J,E}(x) = G_i^{J,E}x + \mu_i^{J,E}$, for all $x \in \Omega$ and $\eta \in \mathbb{R}^n$.

Remark 3: It could be interesting to characterize, in future, the cases for which the existence of affine or linear functions $\sigma_i^{J,E}(x)$ satisfying the hypotheses of Theorem 1 is also necessary for the contractivity of $\alpha\Omega$ in Ω .

Consider the potentially asymmetric saturations given by

$$\varphi_i(y) = \begin{cases} \bar{y}_i & \text{if } y_i > \bar{y}_i, \\ y_i & \text{if } -\underline{y}_i \leq y_i \leq \bar{y}_i, \\ -\underline{y}_i & \text{if } y_i < -\underline{y}_i, \end{cases} \quad (22)$$

with $-\underline{y}_i \leq \bar{y}_i$ for every $i \in \mathbb{N}_m$.

Remark 4: The only requirement is $-\underline{y}_i \leq \bar{y}_i$ for all $i \in \mathbb{N}_m$, no assumption on the saturation levels signs has been posed.

The sets in (4) should be adapted to the new saturations levels. For instance, in case of asymmetric saturations, we have that $\mathcal{N}^+(x, \eta) = \{i \in \mathbb{N}_m : \eta^T B_{(i)} > 0, K_i x < -\underline{y}_i\}$ and $\eta^T B_{(i)} \varphi_i(Kx) = -\eta^T B_{(i)} \underline{y}_i$, if $i \in \mathcal{N}^+(x, \eta)$, see (5). The other definitions, obtainable as direct modifications of (4) and (5), are assumed to be used in the following corollary.

Corollary 1: Given the system (1) with saturations as in (22), the compact convex set $\Omega \subseteq \mathbb{R}^n$, with $0 \in \text{int}(\Omega)$, is such that $\alpha\Omega$ is λ -contractive for every $\alpha \in [0, 1]$ if and only if for every $J \in \mathcal{J}$ and every $E \in \mathcal{K}$, there exist $\gamma_i^{J,E} \in \mathbb{R}$ and

$\sigma_i^{J,E}(x) \in \mathbb{R}$, with $i \in \mathbb{N}_m$, such that condition (8) and

$$\begin{cases} \gamma_i^{J,E} \leq 0, & \sigma_i^{J,E}(x) \geq (\gamma_i^{J,E} - 1)\underline{y}_i, & \text{if } i \in \mathcal{N}^+(J, E), \\ \gamma_i^{J,E} \leq 0, & \sigma_i^{J,E}(x) \leq (1 - \gamma_i^{J,E})\bar{y}_i, & \text{if } i \in \mathcal{N}^-(J, E), \\ \gamma_i^{J,E} \geq 0, & \sigma_i^{J,E}(x) \geq (1 - \gamma_i^{J,E})\bar{y}_i, & \text{if } i \in \mathcal{P}^+(J, E), \\ \gamma_i^{J,E} \geq 0, & \sigma_i^{J,E}(x) \leq (\gamma_i^{J,E} - 1)\underline{y}_i, & \text{if } i \in \mathcal{P}^-(J, E), \\ \gamma_i^{J,E} \in \mathbb{R}, & \begin{cases} \sigma_i^{J,E}(x) \geq (\gamma_i^{J,E} - 1)\underline{y}_i, \\ \sigma_i^{J,E}(x) \geq (1 - \gamma_i^{J,E})\bar{y}_i, \end{cases} & \text{if } i \in \mathcal{L}^+(J, E), \\ \gamma_i^{J,E} \in \mathbb{R}, & \begin{cases} \sigma_i^{J,E}(x) \leq (\gamma_i^{J,E} - 1)\underline{y}_i, \\ \sigma_i^{J,E}(x) \leq (1 - \gamma_i^{J,E})\bar{y}_i, \end{cases} & \text{if } i \in \mathcal{L}^-(J, E), \\ \gamma_i^{J,E} \in \mathbb{R}, & \sigma_i^{J,E}(x) \in \mathbb{R}, & \text{if } i \in \mathcal{O}(E), \end{cases} \quad (23)$$

hold for all $x \in \Omega$ and all $\eta \in \mathbb{R}^n$.

IV. ILLUSTRATIVE EXAMPLE

The main aim of this example is to highlight the substantial improvement provided by our results with respect to similar ones from the literature, see [2], [9]–[12]. For this, we show that, also for trivial one-dimensional systems, the conservativity introduced might be very important. Comparing such approaches with ours, which overcomes the limitations of the formers, illustrates adequately the contribution of the paper.

We stress that our results can apply to more complex functions, e.g. to composite quadratic, as in [10], and polyhedral ones. Nevertheless we focus on quadratic functions, probably the most common and the best known in the invariant studies.

First, we recall the result presented in [2], for the continuous-time context but extendable to the discrete-time one. A first improvement of such a result has appeared in [5].

Theorem 2: Given the system (1), and the ellipsoid $\Omega = \mathcal{E}(P)$, with $P \in \mathbb{R}^{n \times n}$ and $P > 0$, if for every $J \subseteq \mathbb{N}_m$ and every $i \in J$, there exists $G_i^J \in \mathbb{R}^{1 \times n}$ such that

$$\begin{aligned} (N^J)^T P N^J &\leq \lambda P, & \forall J \subseteq \mathbb{N}_m, \\ G_i^J P^{-1} (G_i^J)^T &\leq 1, & \forall J \subseteq \mathbb{N}_m, \forall i \in J, \end{aligned}$$

where $N^J = A + \sum_{i \in J} B_{(i)} K_i + \sum_{i \in J} B_{(i)} G_i^J$, then $\alpha\Omega$ is λ -contractive, with $\lambda \in [0, 1]$, for every $\alpha \in [0, 1]$.

The quadratic stability conditions for saturated systems given in [2], [9]–[12] are substantially based on Theorem 2.

Example 1: Consider the one-dimensional system

$$x^+ = f(x) = x + 0.5\varphi(x) - \varphi(0.5x), \quad (24)$$

that is (1), with $n = 1$, $m = 2$ and $A = 1$, $B = [0.5, -1]$ and $K = [1, 0.5]^T$. An ellipsoid in \mathbb{R} centered in 0 and parameterized with respect to $p > 0$ is $\Omega = \{x \in \mathbb{R} : x^2 p^2 \leq 1\} = \{x \in \mathbb{R} : -p^{-1} \leq x \leq p^{-1}\}$. Consider $\lambda = 1$. The conditions in Theorem 2 are, then, equivalent to the existence of $g_i^J \in \mathbb{R}$, for every $J \subseteq \mathbb{N}_2$ and $i \in J$, such that

$$\begin{aligned} (1 + \sum_{i \in J} B_{(i)} K_i + \sum_{i \in J} B_{(i)} g_i^J)^2 &\leq 1, & \forall J \subseteq \mathbb{N}_2, \\ p^{-2} (g_i^J)^2 &\leq 1, & \forall J \subseteq \mathbb{N}_2, \forall i \in J. \end{aligned} \quad (25)$$

The system is positive and $f(x)$ in (24) is odd symmetric, then we can consider just $x \geq 0$ and $\eta \geq 0$. We have that

$$f(x) = \begin{cases} x, & \text{if } x \in [0, 1), \\ 0.5x + 0.5, & \text{if } x \in [1, 2), \\ x - 0.5, & \text{if } x \in [2, \infty), \end{cases}$$

and either there is no active saturation, or only the first input is saturated or both are saturated. There is not a region of x at which the second input is saturated but not the first one. Nevertheless, all the possible combinations on saturations are implicitly considered in Theorem 2. This is, we guess, an important source of conservatism, overcome by our method.

We consider now all the $J \subseteq \mathbb{N}_m$ and search for the minimal p such that (25) holds. For $J = \emptyset$ we have that $(1 + 0.5 - 0.5) \leq 1$, which holds for all $p \in \mathbb{R}$. Consider $J = \{1\}$, related to the saturation of the first input. The conditions result in

$$(1 + 0.5g_1^{\{1\}} - 0.5) \leq 1, \quad (g_1^{\{1\}})^2 \leq p^2,$$

a solution of which is $g_1^{\{1\}} = 0$ and thus $p = 0$, which implies $\Omega = \mathbb{R}$. Considering both saturations, i.e. $J = \{1, 2\}$, we have

$$(1 + 0.5g_1^{\{1,2\}} - g_2^{\{1,2\}}) \leq 1, \quad (g_1^{\{1,2\}})^2 \leq p^2, \quad (g_2^{\{1,2\}})^2 \leq p^2. \quad (26)$$

Posing $g_1^{\{1,2\}} = g_2^{\{1,2\}} = 0$, conditions (26) hold for all $p \geq 0$ then $\Omega = \mathbb{R}$. Finally we consider $J = \{2\}$ and we have

$$(1 + 0.5 - g_2^{\{2\}}) \leq 1, \quad (g_2^{\{2\}})^2 \leq p^2. \quad (27)$$

For the first constraint of (27) to hold one must have $g_2^{\{2\}} \geq 0.5$, and then, from the second $p \geq 0.5$ which means $\Omega = [2, 2]$. Hence the result of applying Theorem 2 would be ellipsoids $\Omega = \{x \in \mathbb{R} : |x| \leq q\}$ with $q = p^{-1} \leq 2$.

Consider now our approach, namely the result presented in Theorem 1. Given nonnegative x and η we have that $\phi_{\Omega(x)}(\eta) = |\eta||x| = \eta x$, and then condition (8) becomes $\eta(x + \sum_{i \in \mathbb{N}_m} \gamma_i^{J,E} B_{(i)} K_i x + g_i^{J,E}(x) B_{(i)}) \leq \eta x$, or equivalently

$$\gamma_1^{J,E} x + g_1^{J,E}(x) - \gamma_2^{J,E} x - 2g_2^{J,E}(x) \leq 0. \quad (28)$$

First consider the cases of $J \in \mathcal{J}$ and $E \in \mathcal{K}$, i.e. all the combinations except the saturation of the second input only, not admissible. The results presented above, satisfying the conditions of Theorem 2, hold also for those of Theorem 1 with adequate $\gamma_i^{J,E}$. This implies that invariance of $\Omega = \{x \in \mathbb{R} : |x| \leq p^{-1}\}$ is ensured for all $p > 0$. This is the exact solution since $(x, f(x))$ lies in the sector of \mathbb{R}^2 between $(x, 0)$ and (x, x) , for all x , condition for absolute stability to hold.

We prove that the case of saturation of the second input only, not admissible, leads in our approach to negligible constraints, see Remark 2. We have to prove that (8) and (9) can be satisfied for all $x \in \mathbb{R}$ and $\eta \in \mathbb{R}$. Considering $\eta \geq 0$, suppose that $\eta B_{(1)} > 0$ and $|K_1 x| \leq 1$, hence $1 \in \mathcal{L}^+(J, E)$. Since the second input saturates, $K_2 x > 1$ and $\eta B_{(2)} < 0$ then $2 \in \mathcal{N}^-(J, E)$. Thus (9) holds for all $x \in \Omega$ if and only if

$$\begin{cases} g_1^{J,E}(x) \geq \gamma_1^{J,E} - 1, & g_1^{J,E}(x) \geq 1 - \gamma_1^{J,E}, & \gamma_1^{J,E} \in \mathbb{R}, \\ g_2^{J,E}(x) \leq 1 - \gamma_2^{J,E}, & \gamma_2^{J,E} \leq 0, \end{cases} \quad (29)$$

for all $x \in \Omega$. Posing $\gamma_1^{J,E} = 0$ and $\gamma_2^{J,E} = 0$, (28) and (29) are

$$g_1^{J,E}(x) \leq 2g_2^{J,E}(x), \quad g_1^{J,E}(x) \geq 1, \quad g_2^{J,E}(x) \leq 1,$$

which are satisfied with $g_1^{J,E}(x) = 1$ and $g_2^{J,E}(x) = 1$ for all $x \in \mathbb{R}$. Then $\Omega = \{x \in \mathbb{R} : |x| \leq p^{-1}\}$ are invariant for all $p^{-1} \geq 0$. Hence the presence of constraints (8) and (9) related to the inadmissible case (saturation of the second input only) would not affect the result obtained using Theorem 1.

V. CONCLUSION

Necessary and sufficient conditions for a compact convex set to be invariant or contractive for saturated systems have been posed. The characterization of set-induced Lyapunov functions is given. It generalizes in several ways previous works concerning invariance and Lyapunov stability for saturated systems, since it considers asymmetry of sets and saturations and since the conditions are also necessary and apply to generic convex compact sets. Further investigations should deal with the application of these results to continuous-time and hybrid saturated systems, extending the results in [6]. Performances and anti-windup design, based on the presented conditions, may be considered. The computational aspects might also deserve to be addressed.

REFERENCES

- [1] T. Alamo, A. Cepeda, M. Fiacchini, and E. F. Camacho. Convex invariant sets for discrete-time Lur'e systems. *Automatica*, 45:1066–1071, 2009.
- [2] T. Alamo, A. Cepeda, and D. Limon. Improved computation of ellipsoidal invariant sets for saturated control systems. In *Proc. of the 44th IEEE Conf. on Decision and Control and European Control Conf. 2005, CDC-ECC'05*, pages 6216–6221, Seville, Spain, Dec. 2005.
- [3] F. Blanchini and S. Miani. *Set-Theoretic Methods in Control*. Birkhäuser, 2008.
- [4] M. Fiacchini, T. Alamo, and E. F. Camacho. On the computation of convex robust control invariant sets for nonlinear systems. *Automatica*, 46(8):1334–1338, 2010.
- [5] M. Fiacchini, S. Tarbouriech, and C. Prieur. Invariance of symmetric convex sets for discrete-time saturated systems. In *Proc. of the 50th IEEE Conf. on Decision and Control and European Control Conf. 2011, CDC-ECC'11*, pages 7343–7348, Orlando, FL, Dec. 2011.
- [6] M. Fiacchini, S. Tarbouriech, and C. Prieur. Quadratic stability for hybrid systems with nested saturations. *IEEE Transactions on Automatic Control*, 57(7):1832–1838, 2012.
- [7] J. M. Gomes da Silva Jr. and S. Tarbouriech. Polyhedral regions of local stability for linear discrete-time systems with saturating controls. *IEEE Transactions on Automatic Control*, 44:2081–2085, 1999.
- [8] J. M. Gomes Da Silva Jr. and S. Tarbouriech. Local stabilization of discrete-time linear systems with saturating controls: An LMI-based approach. *IEEE Transactions on Automatic Control*, 46:119–125, 2001.
- [9] T. Hu and Z. Lin. Exact characterization of invariant ellipsoids for single input linear systems subject to actuator saturation. *IEEE Transactions on Automatic Control*, 47(1):164–169, 2002.
- [10] T. Hu and Z. Lin. Absolute stability analysis of discrete-time systems with composite quadratic Lyapunov functions. *IEEE Transactions on Automatic Control*, 50(6):781–797, 2005.
- [11] T. Hu, A. R. Teel, and L. Zaccarian. Stability and performance for saturated systems via quadratic and nonquadratic Lyapunov functions. *IEEE Transactions on Automatic Control*, 51(11):1770–1786, 2006.
- [12] T. Hu, T. Thibodeau, and A. R. Teel. A unified Lyapunov approach to analysis of oscillations and stability for systems with piecewise linear elements. *IEEE Transactions on Automatic Control*, 55(12):2864–2869, 2010.
- [13] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, USA, 1970.
- [14] S. Tarbouriech, G. Garcia J. M. Gomes da Silva Jr., and I. Queinnec. *Stability and Stabilization of Linear Systems with Saturating Actuators*. Springer, 2011.
- [15] L. Zaccarian and A. R. Teel. *Modern Anti-windup Synthesis: Control Augmentation for Actuator Saturation*. Princeton Series in Applied Mathematics. Princeton University Press, 2011.
- [16] B. Zhou, W. X. Zheng, and G.-R. Duan. An improved treatment of saturation nonlinearity with its application to control of systems subject to nested saturation. *Automatica*, 47(2):306–315, 2011.