Gradient Projection Anti-windup Scheme on Constrained Planar LTI Systems

Justin Teo and Jonathan P. How

Technical Report ACL10–01
Aerospace Controls Laboratory
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

March 15, 2010

***Abstract*—The *gradient projection anti-windup* (GPAW) scheme was recently proposed as an anti-windup method for *nonlinear* multi-input-multi-output systems/controllers, the solution of which was recognized as a largely open problem in a recent survey paper. This report analyzes the properties of the GPAW scheme applied to an input constrained first order linear time invariant (LTI) system driven by a first order LTI controller, where the objective is to regulate the system state about the origin. We show that the GPAW compensated system is in fact a *projected dynamical system* (PDS), and use results in the PDS literature to assert existence and uniqueness of its solutions. The main result is that the GPAW scheme can only *maintain/enlarge* the *exact* region of attraction of the uncompensated system. We illustrate the qualitative weaknesses of some results in establishing true advantages of anti-windup methods, and propose a new paradigm to address the anti-windup problem, where results *relative* to the uncompensated system are sought.**

***Index Terms*—gradient projection anti-windup, constrained planar LTI systems, projected dynamical systems, equilibria, region of attraction.**

I. Introduction

THE *gradient projection anti-windup* (GPAW) scheme was proposed in [1] as an anti-windup method for *nonlinear* multi-input-multi-output (MIMO) systems/controllers. It was recognized in a recent survey paper [2] that anti-windup compensation for nonlinear systems remains largely an *open problem*. To this end, [3] and relevant references in [2] represent some recent advances. The GPAW scheme uses a continuous-time extension of the gradient projection method of nonlinear programming [4], [5] to extend the "stop integration" heuristic outlined in [6] to the case of nonlinear MIMO systems/controllers. Application of the GPAW scheme to some nominal controllers results in a *hybrid* GPAW compensated controller [1], and hence a hybrid closed loop system.

J. Teo is a graduate student with the Aerospace Controls Laboratory, Department of Aeronautics & Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (email: csteo@mit.edu).

J. How is director of Aerospace Controls Laboratory and Professor in the Department of Aeronautics & Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (email: jhow@mit.edu).

Here, we apply the GPAW scheme to a first order linear time invariant (LTI) system stabilized by a first order LTI controller, where the objective is to regulate the system state about the origin. This case is particularly insightful because the closed loop system is a planar dynamical system whose vector field is easily visualized, and is highly tractable because there is a large body of relevant work, eg. [7, Chapter 2] [8, Chapter 2] [9, Chapter 2] [10, Chapter 3] [11]. Related literature on constrained planar systems include [12]–[18].

After presenting the generalities in Section II, we address the existence and uniqueness of solutions to the GPAW compensated system. Due to *discontinuities* of the governing vector field of the GPAW compensated system on the saturation constraint boundaries, classical existence and uniqueness results based on Lipschitz continuity of vector fields [7]–[10] do not apply directly. We show that the GPAW compensated system is in fact a *projected dynamical system* (PDS) [19]–[22] in Section III. Observe that PDS is a significant line of independent research that has attracted the attention of economists and mathematicians, among others. The link to PDS thus enables cross utilization of ideas and methods, as demonstrated in [23]. Using results from the PDS literature, existence and uniqueness of solutions to the GPAW compensated system can thus be easily established, as shown in Section IV. In Section V, equilibria of the systems are characterized, leading to the study of the associated region of attraction (ROA).

It is widely accepted as a rule that the performance of a control system can be enhanced by trading off its robustness [24, Section 9.1]. As such, we consider an anti-windup scheme to be valid only if it can provide performance enhancements *without reducing the system's ROA*. The first question to be addressed is whether the GPAW scheme satisfy such a criterion, and is shown to be affirmative in Section VI.

Numerical results further illuminate this property of GPAW compensated systems.

In Section VII, we illustrate some qualitative weaknesses of some results in the anti-windup literature, and propose a new paradigm in addressing the anti-windup problem, in which results *relative* to the uncompensated system are sought. This is the case for the main result of this report, Proposition 4.

II. Preliminaries

Let the system to be controlled be described by

$$\dot{x} = ax + b\,\mathrm{sat}(u), \tag{1}$$

where the saturation function is defined by

$$\mathrm{sat}(u) = \begin{cases} u_{max}, & \text{if } u \geq u_{max}, \\ u, & \text{if } u_{min} < u < u_{max}, \\ u_{min}, & \text{if } u \leq u_{min}, \end{cases}$$

and $x, u \in \mathbb{R}$ are the plant state and control input respectively, $a, b, u_{min}, u_{max} \in \mathbb{R}$ are constant plant parameters with u_{min}, u_{max} satisfying $u_{min} < 0 < u_{max}$. Let the *nominal* controller be

$$\begin{aligned} \dot{x}_c &= \tilde{c}x_c + \tilde{d}x, \\ u &= \tilde{e}x_c, \end{aligned} \tag{2}$$

where $x_c, u \in \mathbb{R}$ are the controller state and output respectively, $x \in \mathbb{R}$ is the measurement of the plant state, and $\tilde{c}, \tilde{d}, \tilde{e} \in \mathbb{R}$ are controller gains chosen to *globally* stabilize the *unconstrained* system, ie. when $u_{max} = -u_{min} = \infty$.

Remark 1: It is important that the output equation of the nominal controller, namely $u = \tilde{e}x_c$, depends only on the controller state x_c and independent of measurement x. That is, if the output equation is $u = \tilde{e}x_c + \tilde{f}x$, then we require $\tilde{f} = 0$. This property ensures that full controller state-output consistency, ie. $\mathrm{sat}(u) = u$, can be maintained at "almost all" times (stated more precisely as Fact 1 below) when applying the GPAW scheme. For general nominal controllers, this requirement and its consequences on the GPAW compensated controller are detailed in [25], together with remedies when the nominal controller does not have the required structure. □

A simple transformation of (2) yields the equivalent controller realization

$$\dot{u} = cx + du, \tag{3}$$

with $c := \tilde{d}\tilde{e}$, $d := \tilde{c}$. Applying the GPAW scheme [1] to the preceding transformed nominal controller (3) yields the GPAW compensated controller (see Appendix A)

$$\dot{u} = \begin{cases} 0, & \text{if } u \geq u_{max}, cx + du > 0, \\ 0, & \text{if } u \leq u_{min}, cx + du < 0, \\ cx + du, & \text{otherwise}, \end{cases} \tag{4}$$

which is similar to the "conditionally freeze integrator" method [26]. This is expected since the GPAW scheme can be viewed as a generalization of this idea to MIMO nonlinear controllers. Observe that the first order GPAW compensated controller is independent of the GPAW tuning parameter Γ introduced in [1], which is true for all first order controllers. Furthermore, inspection of (4) reveals the following.

Fact 1 (Controller State-Output Consistency): If for some $T \in \mathbb{R}$, the control signal of the GPAW compensated controller (4) at time T satisfies $u_{min} \leq u(T) \leq u_{max}$, then $u_{min} \leq u(t) \leq u_{max}$ holds for all $t \geq T$. □

That is, the GPAW compensated controller maintains full controller state-output consistency, $\mathrm{sat}(u) = u$, for all future times once it has been achieved for any time instant. In particular, if the controller state is initialized such that $\mathrm{sat}(u(0)) = u(0)$, then $\mathrm{sat}(u(t)) = u(t)$ holds for all $t \geq 0$.

Remark 2: For nonlinear MIMO controllers whose output equation depends *only on the controller state*, the same result (state-output consistency of GPAW compensated controller) holds as shown in [25, Theorem 1]. □

The *nominal* constrained closed-loop system, Σ_n, is described by (1) and (3),

$$\Sigma_n : \begin{cases} \dot{x} = ax + b\,\mathrm{sat}(u), \\ \dot{u} = cx + du, \end{cases}$$

while the GPAW compensated closed-loop system, Σ_g, is described by (1) and (4),

$$\Sigma_g : \begin{cases} \dot{x} = ax + b\,\mathrm{sat}(u), \\ \dot{u} = \begin{cases} 0, & \text{if } u \geq u_{max}, cx + du > 0, \\ 0, & \text{if } u \leq u_{min}, cx + du < 0, \\ cx + du, & \text{otherwise.} \end{cases} \end{cases}$$

Each of these systems can be expressed in the form $\dot{z} = f(z)$ with $f : \mathbb{R}^2 \to \mathbb{R}^2$. The representing functions (vector fields) for systems Σ_n and Σ_g will be denoted by f_n and f_g respectively. The following will be assumed.

Assumption 1: The controller parameters c, d satisfy

$$a + d < 0, \tag{5}$$

$$ad - bc > 0, \tag{6}$$

and $bc \neq 0$. □

The characteristic equation of the *unconstrained* system, ie. Σ_n with $u_{max} = -u_{min} = \infty$, can be verified to be

$$s^2 - (a+d)s + (ad - bc) = 0,$$

so that Assumption 1 ensures that the origin is a globally exponentially stable equilibrium point for the nominal *unconstrained* system. The condition $bc \neq 0$ ensures that c, d can be chosen to satisfy (5) and (6), and that Σ_n is a *feedback* system.

We will need the following sets

$$\begin{aligned}
K &= \{(x,u) \in \mathbb{R}^2 \mid u_{min} < u < u_{max}\},\\
K_+ &= \{(x,u) \in \mathbb{R}^2 \mid u > u_{max}\},\\
K_- &= \{(x,u) \in \mathbb{R}^2 \mid u < u_{min}\},\\
\partial K_+ &= \{(x,u) \in \mathbb{R}^2 \mid u = u_{max}\},\\
\partial K_- &= \{(x,u) \in \mathbb{R}^2 \mid u = u_{min}\},\\
\partial K_{+div} &= \{(x,u) \in \mathbb{R}^2 \mid u > u_{max}, cx + du = 0\},\\
K_{+in} &= \{(x,u) \in \mathbb{R}^2 \mid u > u_{max}, cx + du < 0\},\\
K_{+out} &= \{(x,u) \in \mathbb{R}^2 \mid u > u_{max}, cx + du > 0\},\\
\partial K_{-div} &= \{(x,u) \in \mathbb{R}^2 \mid u < u_{min}, cx + du = 0\},\\
K_{-in} &= \{(x,u) \in \mathbb{R}^2 \mid u < u_{min}, cx + du > 0\},\\
K_{-out} &= \{(x,u) \in \mathbb{R}^2 \mid u < u_{min}, cx + du < 0\},\\
\partial K_{+in} &= \{(x,u) \in \mathbb{R}^2 \mid u = u_{max}, cx + du_{max} < 0\},\\
\partial K_{+out} &= \{(x,u) \in \mathbb{R}^2 \mid u = u_{max}, cx + du_{max} > 0\},\\
\partial K_{-in} &= \{(x,u) \in \mathbb{R}^2 \mid u = u_{min}, cx + du_{min} > 0\},\\
\partial K_{-out} &= \{(x,u) \in \mathbb{R}^2 \mid u = u_{min}, cx + du_{min} < 0\},\\
\bar{K} &= K \cup \partial K_+ \cup \partial K_-,
\end{aligned}$$

and the points

$$z_+ = \left(-\tfrac{d}{c}u_{max}, u_{max}\right), \qquad z_- = \left(-\tfrac{d}{c}u_{min}, u_{min}\right).$$

These sets and associated vector fields are illustrated in Fig. 1 for an open loop *unstable* plant, and in Fig. 2 for an open loop *stable* plant.

Observe that $K_+ = K_{+in} \cup K_{+div} \cup K_{+out}$ and $\partial K_+ = \partial K_{+in} \cup \partial K_{+out} \cup \{z_+\}$, with analogous counterparts for K_- and ∂K_-. Observe further that on ∂K_{+in} and ∂K_{-in}, vector fields of systems Σ_n and Σ_g (f_n and f_g respectively) point into K. On ∂K_{+out}, f_n points into K_+ and f_g points into ∂K_+. On ∂K_{-out}, f_n points into K_- and f_g points into ∂K_-.

By inspection of the vector fields f_n and f_g *from their definitions*, we have the following.

Fact 2: The vector fields f_n and f_g coincide in

$$K \cup K_{+in} \cup K_{-in} \cup \partial K_{+div} \cup \partial K_{-div} \cup \partial K_{+in} \cup \partial K_{-in} \cup \{z_+, z_-\}.$$

That is, they coincide in $\mathbb{R}^2 \setminus (K_{+out} \cup K_{-out} \cup \partial K_{+out} \cup \partial K_{-out})$. □

Fact 3: Any solution of systems Σ_n or Σ_g can pass from K_+ to K if and only if it intersects the line segment ∂K_{+in}, and analogously with respect to K_- and ∂K_{-in}. □

Fact 4: Any solution of system Σ_n can pass from K to K_+ if and only if it intersects the line segment ∂K_{+out}, and analogously with respect to K_- and ∂K_{-out}. □

III. GPAW Compensated Closed Loop System as a Projected Dynamical System

Two of the most fundamental properties required for a meaningful study of dynamic systems is the existence and uniqueness of their solutions. As evident from the definition of the GPAW compensated controller (4), the vector field of the GPAW compensated system, f_g, is in general *discontinuous* on the saturation constraint boundaries ∂K_{+out} ($\subset \partial K_+$) and ∂K_{-out} ($\subset \partial K_-$). Classical results on the existence and uniqueness of solutions [7]–[10] rely on Lipschitz continuity of the governing vector fields, and hence do not apply to GPAW compensated systems. While results in [27] can be used to assert such properties, we will use results from the *projected dynamical system* (PDS) [19]–[22] literature to assert the existence and uniqueness of solutions to GPAW compensated systems. First, we show here that the GPAW compensated system, Σ_g, is in fact a PDS.

Observe that the set $\bar{K}$ is a closed convex set (in fact, a closed convex polyhedron). The interior and boundary of $\bar{K}$ are K and $\partial K_+ \cup \partial K_-$ respectively. Let $P\colon \mathbb{R}^2 \to \bar{K}$ be the projection operator [19] defined for all $y \in \mathbb{R}^2$ by

$$P(y) = \arg\min_{z \in \bar{K}} \|y - z\|,$$

with $\|\cdot\|$ as the Euclidean norm. It can be seen that for any $(x,u) \in \mathbb{R}^2$, $P((x,u)) = (x, \mathrm{sat}(u))$. Next, for any $y \in \bar{K}$, $v \in \mathbb{R}^2$, define the projection of vector v at y by [19], [20]

$$\pi(y,v) = \lim_{\delta \downarrow 0} \frac{P(y + \delta v) - y}{\delta}.$$

Note that the limit is one-sided in the above definition [20]. With f_n being the vector field of Σ_n, written explicitly as

$$f_n(x,u) = \begin{bmatrix} ax + bu \\ cx + du \end{bmatrix}, \qquad \forall (x,u) \in \bar{K},$$

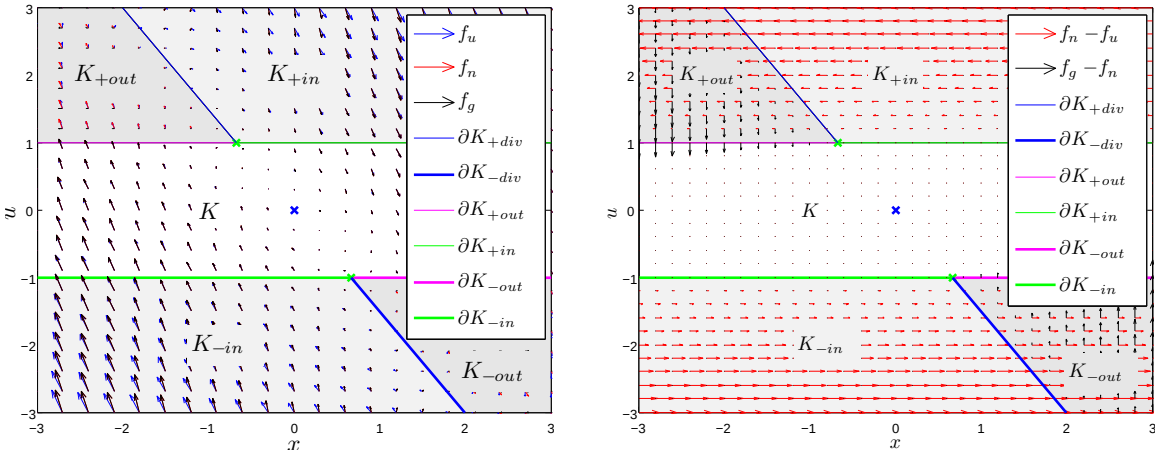


Fig. 1: Closed loop vector fields (f_n, f_g) of systems Σ_n, Σ_g and the *unconstrained* system (Σ_u, f_u), associated with an open loop *unstable* system (plant and controller parameters: $a = 1$, $b = 1$, $c = -3$, $d = -2$, $-u_{min} = u_{max} = 1$). Vector fields of systems Σ_n, Σ_g and Σ_u (f_n, f_g, f_u) are shown on the left, while the vector field differences ($f_n - f_u$, $f_g - f_n$) are shown on the right.

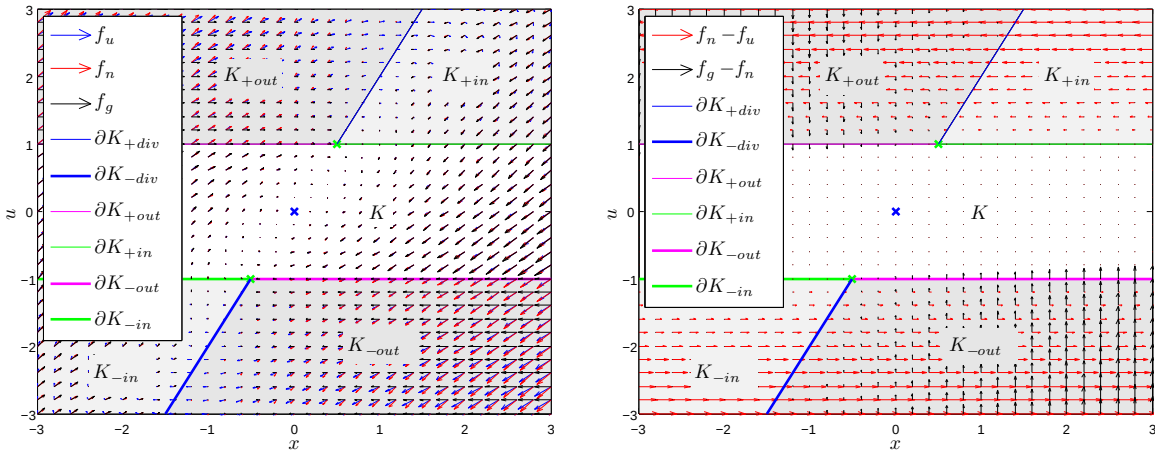


Fig. 2: Closed loop vector fields (f_n, f_g) of systems Σ_n, Σ_g and the *unconstrained* system (Σ_u, f_u) associated with an open loop *stable* system (plant and controller parameters: $a = -1$, $b = 1$, $c = -1$, $d = 0.5$, $-u_{min} = u_{max} = 1$). Vector fields of systems Σ_n, Σ_g and Σ_u (f_n, f_g, f_u) are shown on the left, while the vector field differences ($f_n - f_u$, $f_g - f_n$) are shown on the right.

we have the following, the corollary of which is the desired result.

Claim 1: For all $(x, u) \in \bar{K}$, the vector field f_g of the GPAW compensated closed loop system Σ_g satisfy

$$f_g(x, u) = \pi((x, u), f_n(x, u)).$$

Proof: If $(x, u) \in K$, the result follows from [20, Lemma 2.1(i)] and Fact 2. Next, consider a boundary point, $(x, u) \in \partial K_{+in} \cup \{z_+\}$. On this segment, we have $u = u_{max}$ and $cx + du_{max} \leq 0$ from definition of the set $\partial K_{+in} \cup \{z_+\}$. Since $\text{sat}(u_{max} + \delta\beta) = u_{max} + \delta\beta$ for $\beta \leq 0$ and a sufficiently small $\delta > 0$, we have

$$P((x, u) + \delta f_n(x, u)) = \begin{bmatrix} x + \delta(ax + bu) \\ \text{sat}(u + \delta(cx + du)) \end{bmatrix},$$
$$= \begin{bmatrix} x + \delta(ax + bu) \\ u + \delta(cx + du) \end{bmatrix},$$

so that

$$\pi((x, u), f_n(x, u)) = \lim_{\delta \downarrow 0} \frac{P((x, u) + \delta f_n(x, u)) - (x, u)}{\delta},$$
$$= \begin{bmatrix} ax + bu \\ cx + du \end{bmatrix} = f_n(x, u) = f_g(x, u),$$

for all $(x, u) \in \partial K_{+in} \cup \{z_+\}$, where the final equality follows from Fact 2.

Finally, consider a boundary point $(x,u) \in \partial K_{+out}$. On this segment, we have $u = u_{max}$ and $cx + du_{max} > 0$ from the definition of ∂K_{+out}. Since $\mathrm{sat}(u_{max} + \delta\beta) = u_{max}$ for $\beta > 0$ and a sufficiently small $\delta > 0$, we have

$$\begin{aligned} P((x,u) + \delta f_n(x,u)) &= \begin{bmatrix} x + \delta(ax+bu) \\ \mathrm{sat}(u + \delta(cx+du)) \end{bmatrix}, \\ &= \begin{bmatrix} x + \delta(ax+bu) \\ u \end{bmatrix}, \end{aligned}$$

so that

$$\begin{aligned} \pi((x,u), f_n(x,u)) &= \lim_{\delta \downarrow 0} \frac{P((x,u) + \delta f_n(x,u)) - (x,u)}{\delta}, \\ &= \begin{bmatrix} ax+bu \\ 0 \end{bmatrix} = f_g(x,u), \end{aligned}$$

for all $(x,u) \in \partial K_{+out}$. The above established the claim for all points on $\bar{K} \setminus \partial K_-$. The verification on the boundary ∂K_- is similar to that for ∂K_+. ∎

Corollary 1: The GPAW compensated system Σ_g is a projected dynamical system [19] governed by

$$\dot{z} = f_g(z) = \pi(z, f_n(z)),$$

where $z = (x,u)$.

Corollary 1 will be used in the next section to assert the existence and uniqueness of solutions to system Σ_g. See [19]–[21] for a detailed development of PDS, and [23] for known relations to other system descriptions.

IV. Existence and Uniqueness of Solutions

Here, we assert the existence and uniqueness of solutions to both the nominal constrained system and GPAW compensated system.

Claim 2: The nominal system Σ_n has a unique solution for all initial conditions $(x(t_0), u(t_0)) \in \mathbb{R}^2$ and all $t \geq t_0$.

Proof: For all $z := (x,u) \in \mathbb{R}^2$, the vector field f_n can be written as

$$f_n(z) = Az + \begin{bmatrix} b \\ 0 \end{bmatrix} \mathrm{sat}(u), \qquad \text{where } A = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix}.$$

It can be verified [8, Example 3.2, pp. 91 – 92] that the saturation function is globally Lipschitz with unity Lipschitz constant, ie. $|\mathrm{sat}(\alpha) - \mathrm{sat}(\beta)| \leq |\alpha - \beta|$. Then global Lipschitz continuity of f_n for all $t \geq t_0$ follows from

$$\begin{aligned} \|f_n(z) - f_n(\tilde{z})\| &= \|A(z-\tilde{z}) + [b,0]^{\mathrm{T}}(\mathrm{sat}(u) - \mathrm{sat}(\tilde{u}))\|, \\ &\leq \|A(z-\tilde{z})\| + \|[b,0]^{\mathrm{T}}(\mathrm{sat}(u) - \mathrm{sat}(\tilde{u}))\|, \\ &= \|A(z-\tilde{z})\| + |b||\mathrm{sat}(u) - \mathrm{sat}(\tilde{u})|, \\ &\leq \|A\|\|z-\tilde{z}\| + |b||u-\tilde{u}|, \\ &\leq (\|A\| + |b|)\|z - \tilde{z}\|, \end{aligned} \tag{7}$$

for all $z := (x,u) \in \mathbb{R}^2$, $\tilde{z} := (\tilde{x}, \tilde{u}) \in \mathbb{R}^2$. By [8, Theorem 3.2, pp. 93], Σ_n has a unique solution defined for all $t \geq t_0$, for all $(x(t_0), u(t_0)) \in \mathbb{R}^2$. ∎

We will need the following assumption used to assert the existence and uniqueness of solutions to PDS.

Assumption 2 ([19, Assumption 1]): There exists $B < \infty$ such that the vector field $f_n \colon \mathbb{R}^k \to \mathbb{R}^k$ satisfies the following conditions

$$\|f_n(z)\| \leq B(1 + \|z\|), \qquad \forall z \in \bar{K}, \tag{8}$$

$$\langle f_n(z) - f_n(\tilde{z}), z - \tilde{z} \rangle \leq B\|z - \tilde{z}\|^2, \quad \forall z, \tilde{z} \in \bar{K}, \tag{9}$$

where $\langle x, y \rangle$ denotes the dot product of x and y. □

The following result is stated without proof in the remark following [19, Assumption 1].

Claim 3: If f_n is Lipschitz in $\bar{K} \subset \mathbb{R}^k$, then Assumption 2 holds.

Proof: Since f_n is Lipschitz in $\bar{K}$, there exists an $L < \infty$ such that $\|f_n(z) - f_n(\tilde{z})\| \leq L\|z - \tilde{z}\|$ for all $z, \tilde{z} \in \bar{K}$. To show that (8) holds, observe that

$$\begin{aligned} \|f_n(z)\| &= \|f_n(z) - f_n(\tilde{z}) + f_n(\tilde{z})\|, \\ &\leq \|f_n(z) - f_n(\tilde{z})\| + \|f_n(\tilde{z})\|, \\ &\leq L\|z - \tilde{z}\| + \|f_n(\tilde{z})\|, \\ &\leq L\|z\| + L\|\tilde{z}\| + \|f_n(\tilde{z})\|, \end{aligned}$$

for all $z, \tilde{z} \in \bar{K}$. Fix any $\tilde{z} \in \bar{K}$ and define $\alpha := L\|\tilde{z}\| + \|f_n(\tilde{z})\|$ ($< \infty$) and $B := \max\{L, \alpha\}$ ($< \infty$), so that the preceding inequality becomes

$$\|f_n(z)\| \leq L\|z\| + \alpha \leq B(1 + \|z\|), \qquad \forall z \in \bar{K},$$

which proves (8).

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \langle f_n(z) - f_n(\tilde{z}), z - \tilde{z} \rangle &\leq \|f_n(z) - f_n(\tilde{z})\|\|z - \tilde{z}\|, \\ &\leq L\|z - \tilde{z}\|^2 \leq B\|z - \tilde{z}\|^2, \end{aligned}$$

for all $z, \tilde{z} \in \bar{K}$, which proves (9). ∎

Remark 3: Both Assumption 2 and Claim 3 are stated for general vector fields f_n and regions $\bar{K}$ in $\mathbb{R}^k$, but will be specialized to vector fields and regions in $\mathbb{R}^2$ in the sequel. □

The following is the main result of this section.

Proposition 1: The GPAW compensated system Σ_g has a unique solution for all initial conditions $(x(t_0), u(t_0)) \in \mathbb{R}^2$ and all $t \geq t_0$.

Proof: Since $f_n : \mathbb{R}^2 \to \mathbb{R}^2$ is globally Lipschitz (see (7)), it is Lipschitz in $\bar{K} \subset \mathbb{R}^2$, so that Assumption 2 holds due to Claim 3. Since Σ_g is a PDS (see Corollary 1 and [19, Equation (7)]), it follows from Assumption 2 and [19, Theorem 2] that Σ_g has a unique solution defined for all $t \geq t_0$ whenever the initial condition satisfies $(x(t_0), u(t_0)) \in \bar{K}$ (also recall Fact 1). To assert the existence and uniqueness of solutions for all initial conditions $(x(t_0), u(t_0)) \in \mathbb{R}^2$, it is sufficient to establish this outside $\bar{K}$, and if the solution enters $\bar{K}$, there will be a unique continuation in $\bar{K}$ for all future times from this result.

Consider the region $K_+ = K_{+in} \cup K_{+out} \cup \partial K_{+div}$. The proof for the region K_- is similar. For any $z_1, z_2 \in K_+$, there are three possible cases. Firstly, in the region $\hat{K}_{+out} := K_{+out} \cup \partial K_{+div}$, we get from the definition of f_g and $\hat{K}_{+out}$, that $f_g(z) = f_g(x, u) = (ax + bu_{max}, 0)$. Clearly, for any $z_1, z_2 \in \hat{K}_{+out}$, we have $\|f_g(z_1) - f_g(z_2)\| \leq L_{out}\|z_1 - z_2\|$ where $L_{out} = |a| < \infty$. Secondly, from Fact 2, f_g and f_n coincide in $\hat{K}_{+in} := K_{+in} \cup \partial K_{+div}$, so that f_g is also Lipschitz in $\hat{K}_{+in}$. For any $z_1, z_2 \in \hat{K}_{+in}$, we have $\|f_g(z_1) - f_g(z_2)\| \leq L_{in}\|z_1 - z_2\|$ where $L_{in} = \|A\| + |b| < \infty$ (see (7)). The last case corresponds to z_1 and z_2 being in *different* regions, $\hat{K}_{+in}$ and $\hat{K}_{+out}$. Without loss of generality, let $z_1 \in \hat{K}_{+in}$ and $z_2 \in \hat{K}_{+out}$. The straight line in $\mathbb{R}^2$ connecting z_1 and z_2 then contains a point $\tilde{z} \in \partial K_{+div}$ with the property that $\tilde{z} \in \hat{K}_{+in} \cap \hat{K}_{+out}$, $\|z_1 - \tilde{z}\| \leq \|z_1 - z_2\|$, and $\|z_2 - \tilde{z}\| \leq \|z_1 - z_2\|$. Then we have

$$\begin{aligned}\|f_g(z_1) - f_g(z_2)\| &= \|f_g(z_1) - f_g(\tilde{z}) + f_g(\tilde{z}) - f_g(z_2)\|,\\ &\leq \|f_g(z_1) - f_g(\tilde{z})\| + \|f_g(z_2) - f_g(\tilde{z})\|,\\ &\leq L_{in}\|z_1 - \tilde{z}\| + L_{out}\|z_2 - \tilde{z}\|,\\ &\leq (L_{in} + L_{out})\|z_1 - z_2\|,\end{aligned}$$

which, together with the first two cases, shows that f_g is Lipschitz in K_+. By [9, Theorem 3.1, pp. 18 – 19], Σ_g has a unique solution contained in K_+ whenever $(x(t_0), u(t_0)) \in K_+$. If the solution stays in K_+ for all $t \geq 0$, the claim holds. Otherwise, by [9, Theorem 2.1, pp. 17], the solution can be continued to the boundary of K_+, $\partial K_+ \subset \bar{K}$. In this case, the first part of the proof shows that there is a unique continuation in $\bar{K}$ for all $t \geq 0$. ∎

Remark 4: Care is due when interpreting the existence and uniqueness results of Proposition 1. Let $\phi_n(t, z_0)$ be the unique solution of system Σ_n starting from $z_0 \in \mathbb{R}^2$ at time $t = 0$. For system Σ_n, existence and uniqueness of solution implies that no two different paths intersect [9, pp. 38], and

$$\phi_n(-t, \phi_n(t, z_0)) = z_0, \quad \forall t \in \mathbb{R}, \forall z_0 \in \mathbb{R}^2.$$

That is, proceeding forwards and then backwards in time by the same amount, the solution always reaches its starting point. This is not true for system Σ_g whenever the solution intersects ∂K_{+out} or ∂K_{-out}. Inspection of the vector field f_g reveals that in this case, all *forward* solutions either stay in ∂K_{+out} or ∂K_{-out} for all future times, or they eventually reach the points z_+ or z_-. Furthermore, traversing *backwards in time* from any point of ∂K_{+out} or ∂K_{-out}, the solution stays on these segments indefinitely. That is, ∂K_{+out} and ∂K_{-out} are *negative invariant sets* [9, pp. 47] for system Σ_g. If a *forward* solution of Σ_g intersects ∂K_{+out} or ∂K_{-out} starting from some *interior* point $z_0 \in K$, then traversing backwards in time, the solution will never reach z_0.

Existence and uniqueness of solutions of system Σ_g means that if two distinct trajectories, $\phi_g(t, z_1)$, $\phi_g(t, z_2)$, intersect at some time, then they will be identical for all future times, ie. if $\phi_g(T_1, z_1) = \phi_g(T_2, z_2)$ for some $T_1, T_2 \in \mathbb{R}$, then $\phi_g(t + T_1, z_1) = \phi_g(t + T_2, z_2)$ for all $t \geq 0$. Specifically, they can never diverge into two distinct trajectories. □

V. Equilibrium Points

In this section, we characterize all equilibria of systems Σ_n and Σ_g. Of primary importance is the origin, stated below.

Claim 4: The origin $z_{eq0} := (0, 0)$ is the *only* equilibrium point of systems Σ_n and Σ_g in K, and it must be either a stable node or stable focus.

Proof: In K, the vector fields f_n and f_g coincide (see Fact 2), and can be written as $f_n(z) = f_g(z) = \tilde{A}z$, where $\tilde{A} = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$. It is clear that the origin is an equilibrium point due to $f_n(z_{eq0}) = f_g(z_{eq0}) = \tilde{A}z_{eq0} = 0 \in \mathbb{R}^2$. From (6), the matrix $\tilde{A}$ is invertible and hence, z_{eq0} must be the *only* equilibrium point in K. Due to Assumption 1, z_{eq0} must be either a stable node or a stable focus [7, Section 2.2.1, pp. 32 – 35]. ∎

Additional equilibria of the nominal system Σ_n are characterized below.

Claim 5: Apart from the origin z_{eq0}, the nominal system Σ_n admits two additional isolated equilibrium points defined by

$$z_{eq+} = \left(-\tfrac{b}{a}u_{max}, \tfrac{bc}{ad}u_{max}\right), \quad z_{eq-} = \left(-\tfrac{b}{a}u_{min}, \tfrac{bc}{ad}u_{min}\right),$$

only when

(i) the open loop system is unstable ($a > 0$), or

(ii) the open loop system is *strictly* stable ($a < 0$) and controller parameter satisfies $d \in (0, -a)$.

Moreover, if z_{eq+} and z_{eq-} are equilibria of Σ_n, they are saddle points and lie strictly in K_+ and K_- respectively, ie. $z_{eq+}, z_{eq-} \notin (\partial K_+ \cup \partial K_-)$.

Remark 5: When z_{eq+} and z_{eq-} are equilibria of Σ_n, it can be verified that they must lie in ∂K_{+div} and ∂K_{-div} respectively. □

Proof: All equilibria of Σ_n are determined from the condition $f_n(z) = 0$. It can be verified from the conditions $ax + b\,\mathrm{sat}(u) = 0$ (from $f_n(z) = 0$) and $bc \neq 0$ of Assumption 1, that whenever the open loop system is marginally stable, ie. $a = 0$, there can be no equilibria apart from z_{eq0}. Similarly, whenever $d = 0$, the conditions $cx + du = 0$ (from $f_n(z) = 0$), $bc \neq 0$, and $ax + b\,\mathrm{sat}(u) = 0$ implies that there can be no additional equilibria apart from z_{eq0}. Together, this means $ad \neq 0$, and z_{eq+} and z_{eq-} are well-defined. A simple computation shows that apart from z_{eq0}, the additional equilibria are z_{eq+} and z_{eq-}, provided $z_{eq+} \in K_+ \cup \partial K_+$ and $z_{eq-} \in K_- \cup \partial K_-$. These hold if and only if $ad \neq 0$ and $\frac{bc}{ad} \geq 1$. From (6), $\frac{bc}{ad} \geq 1$ holds if and only if $ad < 0$, which results in the *strict* condition $\frac{bc}{ad} > 1$. Therefore, if z_{eq+} and z_{eq-} are indeed equilibria of Σ_n, they must lie in K_+ and K_- respectively, ie. they cannot lie on ∂K_+ or ∂K_-. If the open loop system is unstable, ie. $a > 0$, then from (5), we must have $d < -a < 0$, which implies $ad < 0$ and Σ_n indeed has z_{eq+} and z_{eq-} as equilibria. If the open loop system is strictly stable, ie. $a < 0$, then $ad < 0$ and (5) hold if and only if $d \in (0, -a)$. It remains to show that z_{eq+} and z_{eq-} must be *saddle points* [7, Section 2.2.1, pp. 32 – 35] whenever they are equilibria of Σ_n.

The Jacobian of f_n at the isolated equilibrium points $z_{eq+} \in K_+$ and $z_{eq-} \in K_-$ are identical and given by

$$\frac{\partial f_n}{\partial z}(z_{eq+}) = \frac{\partial f_n}{\partial z}(z_{eq-}) = A = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix}.$$

Since its eigenvalues are a, d, and $ad < 0$, the equilibria z_{eq+} and z_{eq-} must be saddle points. ■

The following characterizes additional equilibria of the GPAW compensated system Σ_g.

Claim 6: Apart from the origin z_{eq0}, the GPAW compensated system Σ_g admits additional equilibria only when

(i) the open loop system is unstable ($a > 0$). Additional equilibria are all points in the two connected sets defined by

$$Z_{eq+} = \big\{(x,u) \in \mathbb{R}^2 \mid x = -\tfrac{b}{a}u_{max}, \quad u_{max} \leq u \leq \tfrac{bc}{ad}u_{max}\big\} \subset (K_+ \cup \partial K_+),$$
$$Z_{eq-} = \big\{(x,u) \in \mathbb{R}^2 \mid x = -\tfrac{b}{a}u_{min}, \quad \tfrac{bc}{ad}u_{min} \leq u \leq u_{min}\big\} \subset (K_- \cup \partial K_-).$$

(ii) the open loop system is *strictly* stable ($a < 0$) and controller parameter satisfies $d \in (0, -a)$. Additional equilibria are all points in the two connected sets defined by

$$Z_{eq+} = \big\{(x,u) \in \mathbb{R}^2 \mid x = -\tfrac{b}{a}u_{max}, \quad u \geq \tfrac{bc}{ad}u_{max}\big\} \subset K_+,$$
$$Z_{eq-} = \big\{(x,u) \in \mathbb{R}^2 \mid x = -\tfrac{b}{a}u_{min}, \quad u \leq \tfrac{bc}{ad}u_{min}\big\} \subset K_-.$$

Remark 6: Observe that whenever Σ_n has additional equilibria other than z_{eq0}, so does Σ_g. The converse statement is also easily verified. Moreover, observe that z_{eq+} and z_{eq-} belongs to, and lies on the endpoints of the sets Z_{eq+} and Z_{eq-} respectively. □

Proof: All equilibria of Σ_g are determined from the condition $f_g(z) = 0$. It can be verified from the conditions $ax + b\,\mathrm{sat}(u) = 0$ (from $f_g(z) = 0$) and $bc \neq 0$ of Assumption 1, that whenever the open loop system is marginally stable, ie. $a = 0$, there can be no equilibria apart from z_{eq0}. Computation shows that apart from z_{eq0}, all points in the sets

$$Z_{eq+} = \big\{(x,u) \in \mathbb{R}^2 \mid x = -\tfrac{b}{a}u_{max}, \quad u \geq u_{max}, du \geq \tfrac{bc}{a}u_{max}\big\},$$
$$Z_{eq-} = \big\{(x,u) \in \mathbb{R}^2 \mid x = -\tfrac{b}{a}u_{min}, \quad u \leq u_{min}, du \leq \tfrac{bc}{a}u_{min}\big\},$$

are also equilibria of Σ_g, provided these sets are non-empty. Considering the conditions $u \geq u_{max}$ and $du \geq \frac{bc}{a}u_{max}$ (and their analogous counterparts), these sets are non-empty if and only if (a) $d > 0$, (b) $d = 0$ and $\frac{bc}{a} \leq 0$, or (c) $d < 0$ and $\frac{bc}{ad} \geq 1$.

Consider case (a). From (5), this case ($d > 0$) is possible only when $a < 0$, ie. the open loop system is strictly stable. To satisfy (5) and $d > 0$, we must restrict $d \in (0, -a)$. Hence $ad < 0$ and (6) implies $\frac{bc}{ad} > 1$. The above sets Z_{eq+} and Z_{eq-} then simplifies to those stated in the claim for case (ii).

Now consider case (b). With $d = 0$, conditions (5) and (6) reduces to $a < 0$ and $bc < 0$ respectively, which implies $\frac{bc}{a} > 0$. Therefore, Assumption 1 ensures that this case (in particular, $\frac{bc}{a} \leq 0$) cannot occur.

Finally, consider case (c). From (6), this case (in particular, $\frac{bc}{ad} \geq 1$) is possible only when $ad < 0$, which in turn implies

$\frac{bc}{ad} > 1$ holds with *strict* inequality. The condition $ad < 0$ for this case (in particular, $d < 0$) implies $a > 0$, ie. the open loop system is unstable. It is easily verified that the above sets Z_{eq+} and Z_{eq-} simplifies to those stated in the claim for case (i). ■

Remark 7: Observe that the presence of additional equilibria precludes the possibility of the origin being a globally asymptotically stable equilibrium point for both systems Σ_n and Σ_g. However, note that a, d (and b, c) are given fixed parameters in the anti-windup context. □

In summary, z_{eq0} is an isolated stable equilibrium point of systems Σ_n and Σ_g for all $a, b, c, d \in \mathbb{R}$ satisfying Assumption 1, and it is the *only* equilibrium point in K. When the open loop system is marginally stable, or strictly stable with $d \leq 0$, there cannot be additional equilibria. When the open loop system is unstable, or strictly stable with $d \in (0, -a)$, Σ_n has two more isolated equilibrium points z_{eq+} and z_{eq-} which are saddle points, and Σ_g has a continuum of equilibria Z_{eq+} and Z_{eq-}.

VI. Region of Attraction

The purpose of anti-windup schemes is to provide performance improvements only in the presence of control saturation. It is widely accepted as a rule that the performance of a control system can be enhanced by trading off its robustness [24, Section 9.1]. To distinguish anti-windup schemes from conventional control methods, we consider an anti-windup scheme to be valid only if it can provide performance enhancements *without reducing the system's region of attraction (ROA)*. We show in this section that GPAW compensation can only maintain/enlarge the ROA of the nominal system Σ_n. In other words, the ROA of system Σ_n is *contained within* the ROA of Σ_g.

While there may exist multiple equilibria for systems Σ_n and Σ_g, we are primarily interested in the ROA of the equilibrium point at the origin, z_{eq0}. A distinguishing feature is that the results herein refers to the *exact ROA* in contrast to *ROA estimates* that is found in a significant portion of the literature on anti-windup compensation. For clarity of presentation, we present the result in two parts, where the ROA containment is shown for the unsaturated region $\bar{K}$ and saturated region $\mathbb{R}^2 \setminus \bar{K}$ separately. Some numerical examples will illustrate typical ROAs and show that the said ROA containment can hold strictly for some systems. In the sequel, we will state and prove results only for one side of the state space, namely with respect to $K_+ \cup \partial K_+$. The analogous results with respect to $K_- \cup \partial K_-$ can be readily extended, and will not be expressly stated.

Let $\phi_n(t, z_0)$ and $\phi_g(t, z_0)$ be the unique solutions of systems Σ_n and Σ_g respectively, both starting at initial state z_0 at time $t = 0$. The ROA of the origin z_{eq0} for systems Σ_n and Σ_g are then defined by [8, pp. 314]

$$R_n = \{z \in \mathbb{R}^2 \mid \phi_n(t, z) \to z_{eq0} \text{ as } t \to \infty\},$$
$$R_g = \{z \in \mathbb{R}^2 \mid \phi_g(t, z) \to z_{eq0} \text{ as } t \to \infty\},$$

respectively. We recall the notion of *transverse sections* and *ω limit sets*.

Definition 1 ([7, pp. 46]): A *transverse section* σ to a vector field $f \colon \mathbb{R}^2 \to \mathbb{R}^2$ is a continuous, connected arc in $\mathbb{R}^2$ such that the dot product of the unit normal to σ and f is not zero and does not change sign on σ. □

In other words, the vector field has no equilibrium points on σ and is never tangent to σ [7, pp. 46]. It is clear from the definition of ∂K_{+in} and ∂K_{-in} that both of these line segments are in fact transverse sections of f_n and f_g. Moreover, ∂K_{+out} and ∂K_{-out} are also transverse sections of f_n.

Definition 2 ([7, Definition 2.11, pp. 44]): A point $z \in \mathbb{R}^2$ is said to be an *ω limit point* of a trajectory $\phi(t, z_0)$ if there exists a sequence of times $t_n, n \in \{1, 2, \ldots, \infty\}$ such that $t_n \uparrow \infty$ as $n \to \infty$ for which $\lim_{n\to\infty} \phi(t_n, z_0) = z$. The set of all ω limit points of a trajectory is called the *ω limit set* of the trajectory. □

For convenience, let the straight line connecting two points $\alpha, \beta \in \mathbb{R}^2$ be denoted by $l(\alpha, \beta)$ $(= l(\beta, \alpha))$, and defined by

$$l(\alpha, \beta) = \{z \in \mathbb{R}^2 \mid z = \theta\alpha + (1 - \theta)\beta, \forall \theta \in (0, 1)\}.$$

Observe that $l(\alpha, \beta)$ does not contain the endpoints α, β, except for the degenerate case of *identical* endpoints, in which case, $l(\alpha, \alpha) = \{\alpha\}$. Next, the ROA containment in the unsaturated and saturated regions are shown separately, which combines to yield the desired result.

A. ROA Containment in Unsaturated Region

What follows is a series of intermediate claims to arrive at the main result of this subsection, Proposition 2. Let the straight lines connecting the origin to the points z_+ and z_- be

$$\sigma_+ = l(z_{eq0}, z_+) \cup \{z_+\}, \quad \sigma_- = l(z_{eq0}, z_-) \cup \{z_-\}, \quad (10)$$

respectively. Consider a point $z_0 \in \partial K_{+in}$ with the property that $z_0 \in R_n$ and $\phi_n(t, z_0) \notin K_+$ for all $t \geq 0$. In other

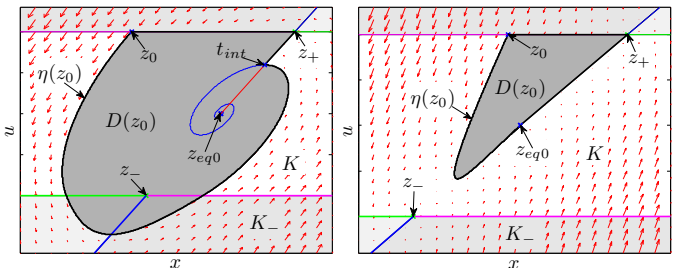


Fig. 3: Closed path $\eta(z_0)$ encloses region $D(z_0) \subset \bar{K} \cup K_-$. A case where the solution enters K_- and also intersects σ_+ is shown on the left, while a case where the solution never enters K_- and never intersects σ_+ is shown on the right.

words, z_0 is in the ROA of system Σ_n and its solution stays in $\bar{K} \cup K_-$ for all $t \geq 0$. As a consequence of Fact 4, $\phi_n(t, z_0)$ can never intersect ∂K_{+out} for all $t \geq 0$. Let

$$t_{int} = \inf\{t \in (0, \infty) \mid \phi_n(t, z_0) \in \sigma_+\}.$$

That is, t_{int} is the first time instant that the solution starting from z_0 at $t = 0$ intersects σ_+, or ∞ if it does not intersect σ_+. If $t_{int} < \infty$, the path

$$\eta_{int}(z_0) = \{z \in \mathbb{R}^2 \mid z = \phi_n(t, z_0), \forall t \in [0, t_{int}]\} \cup l(\phi_n(t_{int}, z_0), z_+) \cup \{z_+\} \cup l(z_0, z_+),$$

is well defined. Otherwise, the path

$$\eta_0(z_0) = \{z \in \mathbb{R}^2 \mid z = \phi_n(t, z_0), \forall t \geq 0\} \cup \{z_{eq0}\} \cup \sigma_+ \cup l(z_0, z_+),$$

is well defined. Now, define the path $\eta(z_0) \in \mathbb{R}^2$ by

$$\eta(z_0) = \begin{cases} \eta_{int}(z_0), & \text{if } t_{int} < \infty, \\ \eta_0(z_0), & \text{otherwise,} \end{cases}$$

which can be verified to be closed and connected. Let the *open, bounded* region enclosed by $\eta(z_0)$ be $D(z_0)$, and its closure be $\bar{D}(z_0)$. The region $D(z_0)$ is illustrated in Fig. 3.

The following result states that $\bar{D}(z_0)$ is a *positive invariant set* [9, pp. 47], and it must contain the origin z_{eq0}.

Claim 7: If there exists a point $z_0 \in \partial K_{+in}$ such that $z_0 \in R_n$ and $\phi_n(t, z_0) \in \bar{K} \cup K_-$ for all $t \geq 0$, then $\bar{D}(z_0) \subset \bar{K} \cup K_-$ is a positive invariant set for system Σ_n, and it must contain z_{eq0}, ie. $z_{eq0} \in \bar{D}(z_0)$.

Remark 8: The claim states specifically that under the assumptions, it is not possible for $\phi_n(t, z_0)$ to intersect σ_+ without having $\eta(z_0)$ enclose z_{eq0}, a case not illustrated in Fig. 3. □

Proof: Let

$$\tilde{\sigma}_+ = \begin{cases} l(\phi_n(t_{int}, z_0), z_+) \cup \{z_+\}, & \text{if } t_{int} < \infty, \\ \sigma_+, & \text{otherwise.} \end{cases}$$

We first show that $\tilde{\sigma}_+$ is a transverse section to f_n, and that f_n always points into $\bar{D}(z_0)$ on $\tilde{\sigma}_+$. Let $\alpha \in \{-1, +1\}$ be chosen such that $\langle \alpha \tilde{T} z_+, z_0 - z_+ \rangle > 0$, where $\tilde{T} z_+ = (u_{max}, \frac{d}{c} u_{max})$ is orthogonal to z_+. Then $\alpha \tilde{T} \frac{z_+}{\|z_+\|}$ is the unit normal of $\tilde{\sigma}_+$ that points into $\bar{D}(z_0)$. Hence $\tilde{\sigma}_+$ is a transverse section to f_n, and f_n points into $\bar{D}(z_0)$ on $\tilde{\sigma}_+$ if and only if $\langle \alpha \tilde{T} z_+, f_n(z) \rangle > 0$ holds with *strict* inequality for all $z \in \tilde{\sigma}_+$.

Since $z_0 \in \partial K_{+in}$, we have from the definition of ∂K_{+in} that $z_0 = (x_0, u_{max})$ for some x_0 that satisfies $c x_0 + d u_{max} < 0$. Then $z_0 - z_+ = (x_0 + \frac{d}{c} u_{max}, 0)$. Due to $c x_0 + d u_{max} < 0$, the condition

$$\begin{aligned} \langle \alpha \tilde{T} z_+, z_0 - z_+ \rangle &= \alpha \left\langle \left(u_{max}, \tfrac{d}{c} u_{max}\right), \left(x_0 + \tfrac{d}{c} u_{max}, 0\right) \right\rangle, \\ &= \tfrac{\alpha}{c} u_{max} (c x_0 + d u_{max}) > 0, \end{aligned}$$

can hold only if $\alpha = -\operatorname{sgn}(c)$. From the definition of $\tilde{\sigma}_+$, any $z \in \tilde{\sigma}_+$ has the form $z = (-\theta \frac{d}{c} u_{max}, \theta u_{max})$ for some $\theta \in (0, 1]$, so that $f_n(z) = ((b - \frac{ad}{c}) \theta u_{max}, 0)$ on $\tilde{\sigma}_+$. Using the definition of f_n on $\tilde{\sigma}_+$, we have

$$\begin{aligned} \langle \alpha \tilde{T} z_+, f_n(z) \rangle &= \alpha \left\langle \left(u_{max}, \tfrac{d}{c} u_{max}\right), \left((b - \tfrac{ad}{c}) \theta u_{max}, 0\right) \right\rangle, \\ &= -\operatorname{sgn}(c) \left(b - \tfrac{ad}{c}\right) \theta u_{max}^2 = \tfrac{ad - bc}{|c|} \theta u_{max}^2. \end{aligned}$$

Since $\theta \in (0, 1]$ for any $z \in \tilde{\sigma}_+$, we have from (6) that $\langle \alpha \tilde{T} z_+, f_n(z) \rangle > 0$, which shows that $\tilde{\sigma}_+$ is a transverse section to f_n and that f_n always points into $\bar{D}(z_0)$ on $\tilde{\sigma}_+$.

It is clear that $l(z_0, z_+) \subset \partial K_{+in}$ is also a transverse section to f_n, and that f_n always points into $\bar{D}(z_0)$ on $l(z_0, z_+)$. Both of these results show that any solution originating in $\bar{D}(z_0)$ cannot exit $\bar{D}(z_0)$ through the line segments $\tilde{\sigma}_+$ or $l(z_0, z_+)$. Furthermore, since the solution is unique and no two different paths can intersect [9, pp. 38], the region $\bar{D}(z_0)$ enclosed by $\eta(z_0)$ must be a *positive invariant set* [9, pp. 47] for system Σ_n. The assumption $\phi_n(t, z_0) \in \bar{K} \cup K_-$ for all $t \geq 0$ implies $\eta(z_0) \subset \bar{K} \cup K_-$. Hence we have $\bar{D}(z_0) \subset \bar{K} \cup K_-$.

Finally, from the assumption $z_0 \in R_n$, we have $\phi_n(t, z_0) \to z_{eq0}$ as $t \to \infty$. Since $\bar{D}(z_0)$ is a positive invariant set and $z_0 \in \bar{D}(z_0)$, we have $\phi_n(t, z_0) \in \bar{D}(z_0)$ for all $t \geq 0$. The conclusion $z_{eq0} \in \bar{D}(z_0)$ then follows from the fact that $\bar{D}(z_0)$ is *closed* and hence contains all its limit points. ■

Claim 8: If there exists a point $z_0 \in \partial K_{+in}$ such that $z_0 \in R_n$ and $\phi_n(t, z_0) \in \bar{K}$ for all $t \geq 0$, then all points in $\bar{D}(z_0) \subset \bar{K}$ also lie in the ROA of system Σ_n, ie. $\bar{D}(z_0) \subset R_n$.

Remark 9: Specifically, the conclusion implies $z_+ \in \bar{D}(z_0) \subset R_n$. □

Proof: Since $\bar{K} \subset (\bar{K} \cup K_-)$, the hypotheses of Claim 7 are satisfied. Claim 7 then shows that $\bar{D}(z_0)$ is a positive invariant set. The condition $\phi_n(t, z_0) \in \bar{K}$ for all $t \geq 0$ implies $\bar{D}(z_0) \subset \bar{K}$. It was shown in [28, Section 6.2, pp. 353 – 363], [9, Theorem 1.3, pp. 55] that for *planar* dynamic systems with only a countable number of equilibria and with unique solutions, the ω limit set of any trajectory contained in any bounded region can only be of three types: equilibrium points, closed orbits, or *heteroclinic/homoclinic orbits* [29, pp. 45], which are unions of saddle points and the trajectories connecting them. It follows from Claims 4 and 5 that the origin z_{eq0} is the *only* equilibrium point of Σ_n in $\bar{K}$, which must be a stable node or stable focus. Hence the ω limit set of any trajectory contained in $\bar{D}(z_0) \subset \bar{K}$ cannot be heteroclinic/homoclinic orbits. By Bendixson's Criterion [8, Lemma 2.2, pp. 67] and (5), region $\bar{D}(z_0)$ contains no closed orbits. As a result, the ω limit sets must consist of equilibrium points only, and it must be z_{eq0} since it is the only equilibrium point in $\bar{K}$. The conclusion follows by observing that $\bar{D}(z_0)$ is a positive invariant set, and any trajectory starting in it must converge to the ω limit set $\{z_{eq0}\}$ due to [8, Lemma 4.1, pp. 127]. ■

The points $\tilde{z}_+ \in \partial K_+$ and $\tilde{z}_- \in \partial K_-$, defined by

$$\tilde{z}_+ := \left(-\tfrac{b}{a}u_{max}, u_{max}\right), \quad \tilde{z}_- := \left(-\tfrac{b}{a}u_{min}, u_{min}\right),$$

and the line segments

$$\xi_+ := l(\tilde{z}_+, z_+) \subset \partial K_+, \quad \xi_- := l(\tilde{z}_-, z_-) \subset \partial K_-,$$

will be needed in the subsequent development.

Claim 9: If the open loop system is stable or marginally stable, ie. $a \leq 0$, then f_g points towards z_+ on ∂K_{+out}, ie. $f_g(z) = \alpha(z_+ - z)$ for some $\alpha = \alpha(z) > 0$ and for all $z \in \partial K_{+out}$. If the open loop system is unstable, ie. $a > 0$, then f_g points towards z_+ on ξ_+, $f_g(\tilde{z}_+) = 0$, and f_g points away from z_+ on $\partial K_{+out} \setminus (\xi_+ \cup \{\tilde{z}_+\})$.

Proof: From the definition of ∂K_{+out}, any $z \in \partial K_{+out}$ has the form $z = (x_0, u_{max})$ for some x_0 satisfying $cx_0 + du_{max} > 0$. For any $z \in \partial K_{+out}$, we have $f_g(z) = (ax_0 + bu_{max}, 0)$ and $z_+ - z = (-(x_0 + \frac{d}{c}u_{max}), 0)$ where $z = (x_0, u_{max})$ and $cx_0 + du_{max} > 0$. The condition $f_g(z) = \alpha(z_+ - z)$ is clearly equivalent to $ax_0 + bu_{max} = -\frac{\alpha}{c}(cx_0 + du_{max})$. Since $cx_0 + du_{max} > 0$, it follows that $f_g(z) = \alpha(z_+ - z)$ can hold with $\alpha > 0$ if and only if

$$c(ax_0 + bu_{max}) < 0. \tag{11}$$

If $a = 0$, (6) reduces to $bc < 0$ and (11) follows. If $a < 0$, we have from (6) and $cx_0 + du_{max} > 0$ that

$$c(ax_0 + bu_{max}) < acx_0 + adu_{max} = a(cx_0 + du_{max}) < 0,$$

and (11) holds. This proves the first statement of the claim.

Finally, consider the case $a > 0$. Then (11) is equivalent to $cx_0 < -\frac{bc}{a}u_{max}$, and $cx_0 + du_{max} > 0$ is equivalent to $cx_0 > -du_{max}$. Hence $f_g(z)$ points towards z_+ on some $z = (x_0, u_{max}) \in \partial K_{+out}$ if and only if x_0 satisfies

$$-du_{max} < cx_0 < -\tfrac{bc}{a}u_{max}. \tag{12}$$

It can be verified that $-du_{max} < -\frac{bc}{a}u_{max}$ due to (6). The above condition (12) can be decomposed and rewritten as

$$\begin{aligned} -\tfrac{d}{c}u_{max} < x_0 < -\tfrac{b}{a}u_{max}, &\qquad \text{if } c > 0, \\ -\tfrac{b}{a}u_{max} < x_0 < -\tfrac{d}{c}u_{max}, &\qquad \text{otherwise,} \end{aligned}$$

so that (12) is equivalent to $x_0 = \left(-\theta\frac{d}{c} - (1-\theta)\frac{b}{a}\right)u_{max}$ for some $\theta \in (0, 1)$. In other words, $f_g(z)$ points towards z_+ if and only if $z \in \xi_+$. The fact that $f_g(\tilde{z}_+) = 0$ can be verified by substitution, and the last statement of the claim follows. ■

Remark 10: It is clear that when $a > 0$, $\tilde{z}_+ \in Z_{eq+}$ where Z_{eq+} is the set of equilibria defined in Claim 6. □

Claim 10: If the open loop system is unstable, ie. $a > 0$, and $z_0 \in \partial K_{+out} \cap R_n$, then $z_0 \in \xi_+$.

Proof: We will show that if $a > 0$ and $z_0 \in \partial K_{+out} \setminus \xi_+$, then $z_0 \notin R_n$ (see Appendix B). If $z_0 \in R_n$, we have $\phi_n(t, z_0) \to z_{eq0}$ as $t \to \infty$. Since $z_{eq0} \in K$, it is sufficient to show that if $a > 0$ and $z_0 \in \partial K_{+out} \setminus \xi_+$, then $\phi_n(t, z_0) \notin K$ for all $t \geq 0$. Let $z_0 = (x_0, u_{max}) \in \partial K_{+out}$ so that $cx_0 + du_{max} > 0$. At the point z_0, we have $f_n(z_0) = (ax_0 + bu_{max}, cx_0 + du_{max})$. It follows that $\dot{u}(0) = cx_0 + du_{max} > 0$ at time $t = 0$, and $u(t)$ must increase (and hence $\mathrm{sat}(u(t)) = u_{max}$) at least for some non-zero interval. The initial value problem to be considered is

$$\begin{aligned} \dot{x} &= ax + bu_{max}, &\qquad x(0) &= x_0, \\ \dot{u} &= cx + du, &\qquad u(0) &= u_{max}, \end{aligned}$$

whose solution will coincide with the solution of Σ_n, ie. $\phi_n(t, z_0)$, as long as it remains outside K. We will show that $u(t) \geq u_{max}$ for all $t \geq 0$, so that $\phi_n(t, z_0) \notin K$ for all $t \geq 0$.

If $c > 0$, we have $-\frac{d}{c}u_{max} < -\frac{b}{a}u_{max}$ from (6). If $z_0 = (x_0, u_{max}) \in \partial K_{+out} \setminus \xi_+$, then x_0 satisfies $x_0 \geq -\frac{b}{a}u_{max}$,

and hence $\dot{x}(0) = ax_0 + bu_{max} \geq 0$. Moreover, because $a > 0$, $x(t)$ is non-decreasing at least until $u(t) < u_{max}$. Hence $x(t) \geq x_0$ and $cx(t) \geq cx_0$ during this interval.

If $c < 0$, then $-\frac{d}{c}u_{max} > -\frac{b}{a}u_{max}$ from (6). If $z_0 = (x_0, u_{max}) \in \partial K_{+out} \setminus \xi_+$, then x_0 satisfies $x_0 \leq -\frac{b}{a}u_{max}$, and hence $\dot{x}(0) = ax_0 + bu_{max} \leq 0$. Moreover, because $a > 0$, $x(t)$ is non-increasing at least until $u(t) < u_{max}$. Hence $x(t) \leq x_0$ and $cx(t) \geq cx_0$ during this interval.

In either case, we have

$$\dot{u} = cx + du \geq cx_0 + du, \qquad u(0) = u_{max},$$

as the differential inequality governing $u(t)$. To apply the Comparison Lemma [8, Lemma 3.4, pp. 102 – 103], define $v = -u$, so that

$$\dot{v} \leq dv - cx_0, \qquad v(0) = -u_{max}.$$

Applying the Comparison Lemma [8, Lemma 3.4, pp. 102 – 103] to the above differential inequality yields $v(t) \leq -u_{max}e^{dt} - \frac{c}{d}x_0\left(e^{dt} - 1\right)$, and hence

$$u(t) = -v(t) \geq u_{max}e^{dt} + \tfrac{c}{d}x_0\left(e^{dt} - 1\right), \quad \forall t \geq 0.$$

Since $a > 0$, it follows from (5) that $d < -a < 0$ and hence $\left(e^{dt} - 1\right) \leq 0$ for all $t \geq 0$. Because $cx_0 + du_{max} > 0$, we have $\frac{c}{d}x_0 < -u_{max}$ and $\frac{c}{d}x_0\left(e^{dt} - 1\right) \geq -u_{max}\left(e^{dt} - 1\right)$. With these, the above inequality becomes

$$\begin{aligned} u(t) &\geq u_{max}e^{dt} + \tfrac{c}{d}x_0\left(e^{dt} - 1\right), \\ &\geq u_{max}e^{dt} - u_{max}\left(e^{dt} - 1\right) = u_{max}, \end{aligned}$$

for all $t \geq 0$. ■

The above results are summarized below.

Claim 11: If there exists a $z_0 \in \partial K_{+out} \cap R_n$, then for every $z \in l(z_0, z_+) \cup \{z_0\}$, there exists a $T(z) \in (0, \infty)$ such that the solution of system Σ_g satisfies $\phi_g(T(z), z) = z_+$ and $\phi_g(t, z) \in \partial K_{+out}$ for all $t \in [0, T(z))$.

Proof: If $a \leq 0$, the result is a direct consequence of Claim 9 and the fact that $\partial K_{+out} \cup \{z_+\}$ contains no equilibrium points of Σ_g. If $a > 0$, then the result follows from Claim 10 and Claim 9, and the fact that $\xi_+ \cup \{z_+\}$ contains no equilibrium points of Σ_g. ■

Remark 11: Observe that under the assumptions, the solution of the GPAW compensated system $\phi_g(t, z_0)$ *slides* along the line segment ∂K_{+out} (or ξ_+ as appropriate) to reach z_+. Note that Fact 1 corroborates this observation. □

Next, we will show that a solution of Σ_n converging to the origin can intersect ∂K_{+out} or ∂K_{-out} only in a specific way, namely that subsequent intersection points, if any, must steadily approach z_+ or z_-.

Claim 12: If $z_0 \in \partial K_{+out} \cap R_n$ and there exists a $T \in (0, \infty)$ such that $\phi_n(T, z_0) \in \partial K_{+out}$, then $\phi_n(T, z_0) \in l(z_0, z_+)$.

Proof: We will show that if $\phi_n(T, z_0) \notin l(z_0, z_+)$, then $z_0 \notin R_n$. Let $z_1 := \phi_n(T, z_0)$ and assume $z_1 \in \partial K_{+out} \setminus l(z_0, z_+)$. If $z_1 = z_0$, then the solution forms a closed orbit, and due to uniqueness of solutions, $\phi_n(t, z_0)$ will stay on the orbit for all $t \geq 0$ and never approach z_{eq0}. Hence $z_0 \notin R_n$. Otherwise, we have $z_1 \in \partial K_{+out} \setminus (l(z_0, z_+) \cup \{z_0\})$. Let the closed bounded region enclosed by the closed path

$$\tilde{\eta}(z_0) = \{z \in \mathbb{R}^2 \mid z = \phi_n(t, z_0), \forall t \in [0, T]\} \cup l(z_0, z_1),$$

be $\tilde{D}(z_0)$. Note that $\phi_n(t, z_0)$ must necessarily intersect ∂K_{+in} and enter K before it can intersect ∂K_{+out} at time T due to Fact 4. It can be seen that $l(z_0, z_1) \subset \partial K_{+out}$ is a transverse section to f_n, with f_n pointing *out* of $\tilde{D}(z_0)$ on $l(z_0, z_1)$. Hence $\tilde{D}(z_0)$ is a *negative invariant set* of system Σ_n. If $z_{eq0} \in \tilde{D}(z_0)$, then there is no way for $\phi_n(t, z_0)$ to reach z_{eq0}, which will prove the claim. We will show that z_{eq0} must be contained in $\tilde{D}(z_0)$ using *index theory* [7, Section 2.4, pp. 49 – 51], [28, Section 5.8, pp. 300 – 305]. Noting that the *index* [7, Definition 2.16, pp. 49] of a closed orbit is $+1$ [28, pp. 301], it can be shown that the index of the closed path $\tilde{\eta}(z_0)$, formed by a section of a trajectory and a transverse section, is also $+1$ [28, pp. 301 – 302]. The indices of a node, focus and saddle are $+1$, $+1$, and -1 respectively [28, pp. 301]. Since the index of $\tilde{\eta}(z_0)$ is the sum of all indices of equilibria enclosed by $\tilde{\eta}(z_0)$ [28, pp. 301], and system Σ_n has only one node or focus at the origin with possibly two additional saddle points, the only way for $\tilde{\eta}(z_0)$ to have an index of $+1$ is for it to enclose the origin z_{eq0} alone. That is, $z_{eq0} \in \tilde{D}(z_0)$. ■

Remark 12: The above proof is most evident by visualizing the vector field f_n on the path $\tilde{\eta}(z_0)$. □

The following is the main result of this subsection. The proof amounts to using the solution of Σ_n to bound the solution of Σ_g.

Proposition 2: The part of the ROA of the origin of system Σ_n contained in $\bar{K}$, is itself contained within the ROA of the origin of system Σ_g, ie. $(R_n \cap \bar{K}) \subset R_g$.

Remark 13: The distinction between the solutions of systems Σ_n and Σ_g, namely $\phi_n(t, z)$ and $\phi_g(t, z)$, and their ROAs, R_n and R_g, should be kept clear when examining the proof below. □

Proof: The following argument will be used repeatedly in

the present proof. If for some $z \in \bar{K}$, we have $\phi_n(t,z) \in \bar{K}$ for all $t \geq 0$, then Fact 4 implies that $\phi_n(t,z)$ cannot intersect ∂K_{+out} or ∂K_{-out}, ie. $\phi_n(t,z) \in \bar{K} \setminus (\partial K_{+out} \cup \partial K_{-out})$ for all $t \geq 0$. Fact 2 shows that f_n and f_g coincide in $\bar{K} \setminus (\partial K_{+out} \cup \partial K_{-out})$, which implies $\phi_g(t,z) = \phi_n(t,z)$ for all $t \geq 0$. If in addition, we have $\lim_{t\to\infty} \phi_n(t,z) = z_{eq0}$, then $\lim_{t\to\infty} \phi_g(t,z) = \lim_{t\to\infty} \phi_n(t,z) = z_{eq0}$. In summary, if $\phi_n(t,z) \in \bar{K}$ for all $t \geq 0$ and $z \in R_n$, then $z \in R_g$. For ease of reference, we call this the *coincidence argument*.

We need to show that if $z_0 \in R_n \cap \bar{K}$, then $z_0 \in R_g$. Let $z_0 \in R_n \cap \bar{K}$, so that $\phi_n(0,z_0) = z_0 \in \bar{K}$, and $\phi_n(t,z_0) \to z_{eq0}$ as $t \to \infty$. Consider the case where $\phi_n(t,z_0)$ stays in $\bar{K}$ for all $t \geq 0$. It follows from the *coincidence argument* that $z_0 \in R_g$.

Now, we let the solution $\phi_n(t,z_0)$ enter K_+ and consider all possible continuations. Due to Fact 4, $\phi_n(t,z_0)$ must intersect ∂K_{+out} at least once. If $\phi_n(t,z_0)$ intersects ∂K_{+out} multiple times, it can only intersect it for finitely many times. Otherwise, there is an infinite sequence of times $t_m, m \in \{1,2,\ldots,\infty\}$ such that $t_m \uparrow \infty$ as $m \to \infty$ for which $\phi_n(t_m,z_0) \in \partial K_{+out}$. Since $z_0 \in R_n$, it follows that $\phi_n(t_m,z_0) \in \partial K_{+out} \cap R_n$ for every m. As a consequence of Claim 12, we have $\lim_{m\to\infty} \phi_n(t_m,z_0) = z_+$, which shows that z_+ is an ω limit point of $\phi_n(t,z_0)$. But this is impossible because $\lim_{t\to\infty} \phi_n(t,z_0) = z_{eq0} \neq z_+$. Similarly, if $\phi_n(t,z_0)$ intersects ∂K_{-out} multiple times, it can only intersect it for finitely many times.

Hence, let T_1 and T_2 be the first and last times for which $\phi_n(t,z_0)$ intersects ∂K_{+out}, and let T_3 be the (*only*) time after T_2 that $\phi_n(t,z_0)$ intersects ∂K_{+in}. Then we have $0 \leq T_1 \leq T_2 < T_3 < \infty$ and $\phi_n(t,z_0) \in K_+$ for all $t \in (T_2,T_3)$, $\phi_n(T_1,z_0), \phi_n(T_2,z_0) \in \partial K_{+out}$, and $\phi_n(T_3,z_0) \in \partial K_{+in}$, with behavior after T_3 to be specified. Let $z_1 = \phi_n(T_1,z_0) \in \partial K_{+out}$, $z_2 = \phi_n(T_2,z_0) \in \partial K_{+out}$ and $z_3 = \phi_n(T_3,z_0) \in \partial K_{+in}$. Since $z_0 \in R_n$, we have $z_1, z_2 \in \partial K_{+out} \cap R_n$ and $z_3 \in \partial K_{+in} \cap R_n$. It is clear that $\phi_g(t,z_0) = \phi_n(t,z_0)$ for all $t \in [0,T_1]$. By Claim 11, there exist a $\tilde{T}_1 < \infty$ such that $\phi_g(T_1+\tilde{T}_1,z_0) = \phi_g(\tilde{T}_1,\phi_g(T_1,z_0)) = \phi_g(\tilde{T}_1,\phi_n(T_1,z_0)) = \phi_g(\tilde{T}_1,z_1) = z_+$. Because $\phi_n(t,z_0)$ cannot intersect ∂K_{+out} for all $t > T_2$, the only possible continuations from time T_3 ($> T_2$) onwards are

(i) $\phi_n(t,z_0)$ stays in $\bar{K}$ for all $t \geq T_3$, or
(ii) $\phi_n(t,z_0)$ enters K_- at some finite time.

Consider case (i), which implies $\bar{D}(z_3) \subset \bar{K}$. Claim 8 yields $z_+ \in \bar{D}(z_3) \subset R_n$, and Claim 7 shows that $\bar{D}(z_3)$ is a positive invariant set for system Σ_n. Then we have $\phi_n(t,z_+) \in \bar{D}(z_3) \subset \bar{K}$ for all $t \geq 0$. It follows from the *coincidence argument* that $z_+ \in R_g$. Because $\phi_g(t,z_+) = \phi_g(t,\phi_g(T_1+\tilde{T}_1,z_0))$ for all $t \geq 0$, we have $z_0 \in R_g$, as desired.

Now, consider case (ii). Due to Fact 4, $\phi_n(t,z_0)$ must intersect ∂K_{-out} at least once. From the above discussion, $\phi_n(t,z_0)$ can intersect ∂K_{-out} only finitely many times. Let T_4 be the first time (after T_3) and T_5 be the last time for which $\phi_n(t,z_0)$ intersects ∂K_{-out}, and let T_6 be the (*only*) time after T_5 that $\phi_n(t,z_0)$ intersects ∂K_{-in}. Then $T_3 < T_4 \leq T_5 < T_6 < \infty$ and $\phi_n(t,z_0) \in K_-$ for all $t \in (T_5,T_6)$, $\phi_n(T_4,z_0), \phi_n(T_5,z_0) \in \partial K_{-out}$, and $\phi_n(T_6,z_0) \in \partial K_{-in}$. Let $z_4 = \phi_n(T_4,z_0) \in \partial K_{-out}$, $z_5 = \phi_n(T_5,z_0) \in \partial K_{-out}$ and $z_6 = \phi_n(T_6,z_0) \in \partial K_{-in}$. Since $z_0 \in R_n$, we have $z_4, z_5 \in \partial K_{-out} \cap R_n$ and $z_6 \in \partial K_{-in} \cap R_n$. Now, the only possible continuation after T_6 is for $\phi_n(t,z_0) \in \bar{K}$ for all $t \geq T_6$. Recall the definition of $\eta(z)$ and $\bar{D}(z)$ for some $z \in \partial K_{+in} \cap R_n$, as illustrated in Fig. 3. It is clear that $z_+ \in \bar{D}(z_3)$. Claim 7 shows that $\bar{D}(z_3)$ (with a portion in K_-) is a positive invariant set for system Σ_n, so that $\phi_n(t,z_+) \in \bar{D}(z_3)$ for all $t \geq 0$. Recall also, that $\phi_g(T_1+\tilde{T}_1,z_0) = z_+$ and we want to show that $z_+ \in R_g$. There are two possible ways for the solution $\phi_n(t,z_+)$ to continue. Either $\phi_n(t,z_+)$ stays in $\bar{D}(z_3) \cap \bar{K}$ for all $t \geq 0$, or it enters $\bar{D}(z_3) \cap K_-$ at some finite time. If $\phi_n(t,z_+) \in \bar{D}(z_3) \cap \bar{K}$ for all $t \geq 0$, then as in the proof of Claim 8, Bendixson's Criterion [8, Lemma 2.2, pp. 67] and the absence of saddle points in $\bar{D}(z_3) \cap \bar{K}$ means that $\{z_{eq0}\}$ is the ω limit set of $\phi_n(t,z_+)$ and hence $z_+ \in R_n$. By the *coincidence argument*, we have $z_+ \in R_g$. It follows from $\phi_g(t,z_+) = \phi_g(t,\phi_g(T_1+\tilde{T}_1,z_0))$ for all $t \geq 0$, that $z_0 \in R_g$. Finally, consider when $\phi_n(t,z_+)$ enters $\bar{D}(z_3) \cap K_-$ at some finite time. By Fact 4, $\phi_n(t,z_+)$ must intersect ∂K_{-out} at least once. Let $\tilde{T}_2 < \infty$ be such that $\phi_n(\tilde{T}_2,z_+) \in \partial K_{-out}$ and $\phi_n(t,z_+) \in K$ for all $t \in (0,\tilde{T}_2)$, and let $\tilde{z}_2 = \phi_n(\tilde{T}_2,z_+) \in \partial K_{-out}$. Because the boundary of $\bar{D}(z_3)$ intersects ∂K_{-out} at z_4 and $\tilde{z}_2 \in \bar{D}(z_3) \cap \partial K_{-out}$, we have that $\tilde{z}_2 \in l(z_4,z_-)$. Since $z_4 \in \partial K_{-out} \cap R_n$, we have by (the analogous counterpart to) Claim 11 that there exists a $\tilde{T}_3 < \infty$ such that $\phi_g(\tilde{T}_3,\tilde{z}_2) = z_-$. Since $z_6 \in \partial K_{-in} \cap R_n$, it follows from (the analogous counterparts to) Claims 8 and 7 that $z_- \in \bar{D}(z_6) \subset R_n$, $\bar{D}(z_6)$ is a positive invariant set, and $\phi_n(t,z_-) \in \bar{D}(z_6) \subset \bar{K}$ for all $t \geq 0$. The *coincidence argument* then yields $z_- \in R_g$. Since $\phi_n(t,z_+) \in K \cup \{z_+\}$ for all $t \in [0,\tilde{T}_2)$, Fact 2 implies that $\phi_g(t,z_+) = \phi_n(t,z_+)$ for all $t \in [0,\tilde{T}_2]$. We can trace back the path to z_0 by observing that $\phi_g(t,z_-) = \phi_g(t,\phi_g(\tilde{T}_3,\tilde{z}_2)) = \phi_g(t+$

$\tilde{T}_3, \tilde{z}_2) = \phi_g(t + \tilde{T}_3, \phi_n(\tilde{T}_2, z_+)) = \phi_g(t + \tilde{T}_3, \phi_g(\tilde{T}_2, z_+)) = \phi_g(t + \tilde{T}_3 + \tilde{T}_2, z_+) = \phi_g(t + \tilde{T}_3 + \tilde{T}_2, \phi_g(T_1 + \tilde{T}_1, z_0))$ for all $t \geq 0$. Since $z_- \in R_g$, we have $z_0 \in R_g$, as desired.

In similar manner, it can be shown that if $z_0 \in R_n \cap \bar{K}$ and the solution $\phi_n(t, z_0)$ enters K_- first, then $z_0 \in R_g$. ■

Observe that the *partial result* stated in Proposition 2 is practically meaningful because the controller state can usually be initialized in a manner such that the system state is in the unsaturated region.

B. ROA Containment in Saturated Region

In this subsection, we show that the ROA containment also holds in the saturated region. What follows is a series of intermediate claims to arrive at the main result of this subsection, Proposition 3.

Define the line segments

$$\sigma_{+div} := \partial K_{+div} \cap \left\{(x,u) \in \mathbb{R}^2 \mid u < \tfrac{bc}{ad} u_{max}\right\},$$
$$\sigma_{-div} := \partial K_{-div} \cap \left\{(x,u) \in \mathbb{R}^2 \mid u > \tfrac{bc}{ad} u_{min}\right\},$$
$$\tilde{\sigma}_{+div} := \partial K_{+div} \setminus \sigma_{+div}, \quad \tilde{\sigma}_{-div} := \partial K_{-div} \setminus \sigma_{-div}.$$

It can be verified that $\sigma_{+div} = l(z_+, z_{eq+})$, $\sigma_{-div} = l(z_-, z_{eq-})$, $z_{eq+} \in \tilde{\sigma}_{+div}$ and $z_{eq-} \in \tilde{\sigma}_{-div}$ whenever $ad < 0$.

Claim 13: If the open loop system is

(i) marginally stable ($a = 0$), or strictly stable with a stable controller ($a < 0$ and $d \leq 0$), then ∂K_{+div} is a transverse section to f_n.

(ii) strictly stable with an unstable controller ($a < 0$ and $d \in (0, -a)$), or unstable ($a > 0$), then σ_{+div} ($\subset \partial K_{+div}$) is a transverse section to f_n.

Proof: Since $\sigma_{+div} \subset \partial K_{+div} \subset K_+$, we only need to consider f_n in K_+. For any $z \in K_+$, we have $f_n(z) = f_n(x,u) = (ax + bu_{max}, cx + du)$. Let $\tilde{T} z_+ = (u_{max}, \frac{d}{c} u_{max})$. For case (i) (respectively, (ii)), it can be verified that $\tilde{T} \frac{z_+}{\|z_+\|}$ is a unit normal of ∂K_{+div} (respectively, σ_{+div}). We need to show that $\langle \tilde{T} z_+, f_n(z) \rangle \neq 0$ for all $z \in \partial K_{+div}$ (respectively, $z \in \sigma_{+div}$). Any $z \in \partial K_{+div}$ can be expressed as $z = (x,u) = (-\frac{d}{c}u, u)$ for some $u > u_{max}$. On any point $z \in \partial K_{+div}$, direct computation yields

$$\begin{aligned}
\langle \tilde{T} z_+, f_n(z) \rangle &= \left\langle \left(u_{max}, \tfrac{d}{c} u_{max}\right), (ax + bu_{max}, cx + du) \right\rangle, \\
&= \left\langle \left(u_{max}, \tfrac{d}{c} u_{max}\right), \left(-\tfrac{ad}{c} u + bu_{max}, 0\right) \right\rangle, \\
&= \left(-\tfrac{ad}{c} u + bu_{max}\right) u_{max}, \\
&= -\tfrac{1}{c}(adu - bcu_{max}) u_{max},
\end{aligned}$$

for some $u > u_{max}$.

For case (i), we have $ad \geq 0$, so that $adu \geq adu_{max} > bcu_{max}$, where the last inequality is due to (6). Then $adu - bcu_{max} > 0$, and we have $\langle \tilde{T} z_+, f_n(z) \rangle \neq 0$ for all $z \in \partial K_{+div}$, as desired.

For case (ii), it can be verified that $ad < 0$ due in part to (5) ($d < -a < 0$ when $a > 0$). Then $\frac{bc}{ad} > 1$ due to (6) and $\sigma_{+div} \neq \emptyset$. On ∂K_{+div}, $\langle \tilde{T} z_+, f_n(z) \rangle = 0$ can hold if and only if $adu - bcu_{max} = 0$. This is assured on any point $z = (x,u) \in \sigma_{+div} \subset \partial K_{+div}$ due to $u < \frac{bc}{ad} u_{max}$. ■

Remark 14: For case (ii), the proof also shows that $\tilde{\sigma}_{+div} \setminus \{z_{eq+}\}$ is also a transverse section to f_n. □

Claim 14: If the open loop system is

(i) strictly stable with an unstable controller ($a < 0$ and $d \in (0, -a)$), or

(ii) unstable ($a > 0$),

and $z_0 \in R_n$, then $z_0 \notin \tilde{\sigma}_{+div}$.

Proof: We will show that if $z_0 \in \tilde{\sigma}_{+div}$, then $z_0 \notin R_n$. If $z_0 \in R_n$, we have $\phi_n(t, z_0) \to z_{eq0}$ as $t \to \infty$. Since $z_{eq0} \in K$, it is sufficient to show that if $z_0 \in \tilde{\sigma}_{+div}$, then $\phi_n(t, z_0) \notin K$ for all $t \geq 0$. It can be verified that $ad < 0$, due in part to (5) ($d < -a < 0$ when $a > 0$). Then (6) yields $\frac{bc}{ad} > 1$. Let $z_0 = (x_0, u_0) \in \tilde{\sigma}_{+div}$, so that $u_0 \geq \frac{bc}{ad} u_{max} > u_{max}$ and $cx_0 + du_0 = 0$. Since $z_0 \in \tilde{\sigma}_{+div} \subset K_+$, consider the initial value problem

$$\begin{aligned}
\dot{x} &= ax + bu_{max}, & x(0) &= x_0, \\
\dot{u} &= cx + du, & u(0) &= u_0,
\end{aligned}$$

whose solution will coincide with $\phi_n(t, z_0)$ as long as it remains in $K_+ \cup \partial K_+$. Solving for $x(t)$ yields

$$x(t) = x_0 e^{at} + \tfrac{b}{a} u_{max}(e^{at} - 1), \qquad \forall t \geq 0.$$

We will show that $u(t) \geq u_{max}$ for all $t \geq 0$, so that $\phi_n(t, z_0) \notin K$ for all $t \geq 0$.

Consider case (i) ($a < 0$ and $d \in (0, -a)$). Define $v = u - u_0$ so that $\dot{v} = \dot{u} = cx + du = dv + (cx + du_0)$, and consider

$$\dot{v} = dv + (cx + du_0), \qquad v(0) = u(0) - u_0 = 0.$$

Clearly, if $v(t) = u(t) - u_0 \geq 0$ for all $t \geq 0$, then $u(t) \geq u_0 \geq \frac{bc}{ad} u_{max} > u_{max}$ for all $t \geq 0$, and the conclusion follows. Since $d > 0$, a sufficient condition is for the input of the preceding ordinary differential equation to satisfy $cx(t) + du_0 \geq 0$ for all $t \geq 0$. Using $cx_0 + du_0 = 0$ and the solution of $x(t)$, we have

$$cx(t) + du_0 = cx_0 e^{at} + \tfrac{bc}{a} u_{max}(e^{at} - 1) + du_0,$$

$$
\begin{aligned}
&= -du_0e^{at} + \tfrac{bc}{a}u_{max}(e^{at}-1) + du_0,\\
&= \left(du_0 - \tfrac{bc}{a}u_{max}\right)(1-e^{at}),\\
&= d\left(u_0 - \tfrac{bc}{ad}u_{max}\right)(1-e^{at}) \geq 0,
\end{aligned}
$$

for all $t \geq 0$, where the final inequality is due to $a < 0$, $d > 0$ and $u_0 \geq \frac{bc}{ad}u_{max}$.

Now consider case (ii) ($a > 0$). From (5), we have $d < -a < 0$. In turn, we have $\frac{bc}{a}u_{max} \geq du_0$ due to $u_0 \geq \frac{bc}{ad}u_{max}$. Because $a > 0$, we have $e^{at} - 1 \geq 0$ for all $t \geq 0$. The evolution of $cx(t)$ then satisfy

$$
\begin{aligned}
cx(t) &= c(x_0e^{at} + \tfrac{b}{a}u_{max}(e^{at}-1)),\\
&\geq cx_0e^{at} + du_0(e^{at}-1) = -du_0 = cx_0,
\end{aligned}
$$

for all $t \geq 0$, due to $cx_0 + du_0 = 0$. Then $u(t)$ is governed by the differential inequality

$$
\dot{u} = cx + du \geq cx_0 + du, \qquad u(0) = u_0.
$$

In similar manner as the proof of Claim 10, define $\tilde{v} = -u$, so that

$$
\dot{\tilde{v}} \leq d\tilde{v} - cx_0, \qquad \tilde{v}(0) = -u_0.
$$

Applying the Comparison Lemma [8, Lemma 3.4, pp. 102 – 103] to the above differential inequality yields $\tilde{v}(t) \leq -u_0e^{dt} - \frac{c}{d}x_0\left(e^{dt}-1\right)$. Using $cx_0 + du_0 = 0$, we have

$$
\begin{aligned}
u(t) = -\tilde{v}(t) &\geq u_0e^{dt} + \tfrac{c}{d}x_0\left(e^{dt}-1\right),\\
&= \tfrac{1}{d}\left(du_0e^{dt} + cx_0\left(e^{dt}-1\right)\right),\\
&= -\tfrac{c}{d}x_0 = u_0 > u_{max},
\end{aligned}
$$

for all $t \geq 0$. ∎

Claim 15: If $z_0 \in K_{+out} \cap R_n$, then there exists a $T_g \in (0,\infty)$ such that the solution of the GPAW compensated system satisfy $\phi_g(T_g, z_0) \in l(z_+, \phi_n(T_n, z_0)) \subset \partial K_{+in}$, where $T_n \in (0,\infty)$ is such that the solution of the nominal system satisfy $\phi_n(T_n, z_0) \in \partial K_{+in}$ and $\phi_n(t, z_0) \in K_+$ for all $t \in [0, T_n)$. Moreover, $T_g < T_n < \infty$ holds.

Proof: Let $z_0 = (x_0, u_0) \in K_{+out} \cap R_n$, so that $u_0 > u_{max}$, $cx_0 + du_0 > 0$, and $\phi_n(t, z_0) \to z_{eq0}$ as $t \to \infty$. Since $z_{eq0} \in K$ and $z_0 \in K_{+out} \subset K_+$, Fact 3 shows that $\phi_n(t, z_0)$ must intersect ∂K_{+in} at some finite time. Let T_n be the first time instant that $\phi_n(t, z_0)$ intersects ∂K_{+in}. Then $T_n \in (0,\infty)$, $\phi_n(T_n, z_0) \in \partial K_{+in}$ and $\phi_n(t, z_0) \in K_+$ for all $t \in [0, T_n)$. It is clear that $l(z_+, \phi_n(T_n, z_0)) \subset \partial K_{+in}$.

The solution of the nominal system, $\phi_n(t, z_0) = (x_n(t), u_n(t))$, is governed by

$$
\begin{aligned}
\dot{x}_n &= ax_n + bu_{max}, \qquad & x_n(0) &= x_0,\\
\dot{u}_n &= h_n(x_n, u_n) = cx_n + du_n, \qquad & u_n(0) &= u_0,
\end{aligned}
$$

as long as $u_n(t) \geq u_{max}$, ie. for all $t \leq T_n$. The solution of the GPAW compensated system, $\phi_g(t, z_0) = (x_g(t), u_g(t))$, is governed by

$$
\begin{aligned}
\dot{x}_g &= ax_g + bu_{max}, \qquad & x_g(0) &= x_0,\\
\dot{u}_g &= h_g(x_g, u_g), \qquad & u_g(0) &= u_0.
\end{aligned}
$$

where

$$
h_g(x_g, u_g) = \begin{cases} 0, & \text{if } cx_g + du_g \geq 0,\\ cx_g + du_g, & \text{otherwise,} \end{cases}
$$

as long as $u_g(t) \geq u_{max}$. We need to show that there exists a $T_g \in (0,\infty)$ such that $T_g < T_n$ and $\phi_g(T_g, z_0) \in l(z_+, \phi_n(T_n, z_0))$.

Solving the initial value problem

$$
\dot{x} = ax + bu_{max}, \quad x(0) = x_0,
$$

yields

$$
x(t) = \begin{cases} x_0 + bu_{max}t, & \text{if } a = 0,\\ x_0e^{at} + \frac{b}{a}u_{max}(e^{at}-1), & \text{otherwise,} \end{cases} \tag{13}
$$

for all $t \geq 0$. It can be seen that $x_n(t) = x(t)$ for all t such that $u_n(t) \geq u_{max}$, and $x_g(t) = x(t)$ for all t such that $u_g(t) \geq u_{max}$. Let $T_g = \inf\{t \in (0,\infty) \mid u_g(t) \leq u_{max}\}$, ie. T_g is the first time instant that $\phi_g(t, z_0)$ intersects ∂K_{+in}, or ∞ if $\phi_g(t, z_0)$ never intersects ∂K_{+in}. With $T := \min\{T_n, T_g\}$, the preceding relations yield $x_n(t) = x_g(t) = x(t)$ for all $t \in [0, T]$. Hence $x_n(t)$ and $x_g(t)$ are well defined at least for all $t \in [0, T]$, and we have

$$
\begin{aligned}
h_n(t, u_n(t)) &:= h_n(x_n(t), u_n(t)) = cx(t) + du_n(t),\\
h_g(t, u_g(t)) &:= h_g(x_g(t), u_g(t))\\
&= \begin{cases} 0, & \text{if } cx(t) + du_g(t) \geq 0,\\ cx(t) + du_g(t), & \text{otherwise.} \end{cases}
\end{aligned}
$$

Observe that whenever $cx(t) + du_g(t) \geq 0$, we have $h_n(t, u_g(t)) \geq h_g(t, u_g(t))$. When $cx(t) + du_g(t) < 0$, we have $h_n(t, u_g(t)) = h_g(t, u_g(t))$. Hence $h_g(t, u_g(t)) \leq h_n(t, u_g(t))$ for all $u_g(t) \geq u_{max}$, for all $t \in [0, T]$. The solution of $u_g(t)$ is clearly governed by the differential inequality

$$
\dot{u}_g(t) = h_g(t, u_g(t)) \leq h_n(t, u_g(t)), \qquad u_g(0) = u_0,
$$

for all $t \in [0, T]$. By the Comparison Lemma [8, Lemma 3.4, pp. 102 – 103], we have $u_g(t) \leq u_n(t)$ for all $t \in [0, T]$.

To obtain a *strict* inequality, observe that $cx(0) + du_n(0) =$

$cx(0) + du_g(0) = cx_0 + du_0 > 0$ holds with strict inequality. Then for any sufficiently small $\delta > 0$ such that $cx(t)+du_n(t) > 0$, $cx(t)+du_g(t) > 0$, for all $t \in [0, \delta]$, there exists an $\epsilon = \epsilon(\delta) > 0$ such that $u_n(\delta) = u_0 + \epsilon$. Moreover, we have $u_g(\delta) = u_0$ due to $\dot{u}_g(t) = 0$ for all $t \in [0, \delta]$. In other words, defining $u_\delta := u_0 + \epsilon$, we have

$$\begin{aligned} \dot{u}_n(t) &= h_n(t, u_n(t)), & u_n(\delta) &= u_\delta, \\ \dot{u}_g(t) &\le h_n(t, u_g(t)), & u_g(\delta) &< u_\delta. \end{aligned}$$

Applying Lemma 1 (see Appendix C) to the preceding, we get the *strict* condition $u_g(t) < u_n(t)$ for all $t \in [\delta, T]$. Since $\delta > 0$ is only required to be small but otherwise arbitrary, we have $u_g(t) < u_n(t)$ for all $t \in (0, T]$.

Assume for the sake of contradiction that $\phi_g(t, z_0)$ never intersects ∂K_{+in}. Then $T_g = \infty$ by its definition and $T := \min\{T_n, T_g\} = T_n < \infty$. Since $\phi_n(T_n, z_0) \in \partial K_{+in}$, we have $u_n(T_n) = u_{max}$. The condition $u_g(t) < u_n(t)$ for all $t \in (0, T_n]$ yields $u_g(T_n) < u_n(T_n) = u_{max}$. This, coupled with $u_g(0) = u_0 > u_{max}$ and continuity of $u_g(t)$ means that there exists a $\tilde{T} \in (0, T_n)$ such that $u_g(\tilde{T}) = u_{max}$. This contradicts the assumption that $\phi_g(t, z_0)$ never intersects ∂K_{+in}, and also shows that $T_g = \tilde{T} < T_n < \infty$ and $\phi_g(T_g, z_0) \in \partial K_{+in}$. It remains to show that $\phi_g(T_g, z_0) \in l(z_+, \phi_n(T_n, z_0))$.

Since $z_0 \in K_{+out}$ and $\phi_n(T_n, z_0), \phi_g(T_g, z_0) \in \partial K_{+in}$, both $\phi_n(t, z_0)$ and $\phi_g(t, z_0)$ must intersect ∂K_{+div} at least once. Let $T_{ndiv} \in (0, T_n)$ and $T_{gdiv} \in (0, T_g)$ be the first time instants that $\phi_n(t, z_0)$ and $\phi_g(t, z_0)$ intersect ∂K_{+div}, so that $\phi_n(t, z_0) \in K_{+out}$ for all $t \in [0, T_{ndiv})$ and $\phi_g(t, z_0) \in K_{+out}$ for all $t \in [0, T_{gdiv})$. This implies that $cx(t)+du_g(t) > 0$ for all $t \in [0, T_{gdiv})$, so that $\dot{u}_g(t) = 0$ and $u_g(t) = u_0$ for all $t \in [0, T_{gdiv}]$. It also implies that $\dot{u}_n(t) = cx(t) + du_n(t) > 0$ for all $t \in [0, T_{ndiv})$, so that $u_n(T_{ndiv}) > u_0$. Then we have $u_{max} < u_0 = u_g(T_{gdiv}) < u_n(T_{ndiv})$, which implies $\phi_g(T_{gdiv}, z_0) \in l(z_+, \phi_n(T_{ndiv}, z_0)) \subset \partial K_{+div}$. Let $z_n := \phi_n(T_{ndiv}, z_0)$, $z_g := \phi_g(T_{gdiv}, z_0)$, and let the closed bounded region enclosed by the closed path

$$\tilde{\eta}(z_n) := l(z_+, z_n) \cup \tilde{\eta}_{\phi_n}(z_n) \cup l(z_+, \phi_n(T_n, z_0)) \cup \{z_+\},$$

where

$$\tilde{\eta}_{\phi_n}(z_n) = \{z \in \mathbb{R}^2 \mid z = \phi_n(t, z_0), \forall t \in [T_{ndiv}, T_n]\},$$

be $\tilde{D}(z_n)$.

If the open loop system is marginally stable ($a = 0$) or strictly stable with a stable controller ($a < 0$ and $d \le 0$), Claim 13 shows that ∂K_{+div} is a transverse section to f_n. Since $\phi_n(t, z_0)$ traverses from K_{+out} through $z_n \in \partial K_{+div}$ to K_{+in}, all trajectories of Σ_n intersecting the transverse section ∂K_{+div} can only pass from K_{+out} to K_{+in}, ie. they cannot pass from K_{+in} to K_{+out} through ∂K_{+div}. This implies that $\phi_n(t, z_0)$ can never return to K_{+out} within the interval $[T_{ndiv}, T_n]$, and $\tilde{D}(z_n)$ is contained in $K_{+in} \cup \partial K_{+div} \cup \partial K_{+in} \cup \{z_+\}$. Moreover, $l(z_+, z_n) \subset \partial K_{+div}$ is also a transverse section to f_n.

If the open loop system is strictly stable with an unstable controller ($a < 0$ and $d \in (0, -a)$), or unstable ($a > 0$), the assumption $z_0 \in R_n$ and Claim 14 implies $z_n \in \sigma_{+div}$, and that $\phi_n(t, z_0) \notin \tilde{\sigma}_{+div}$ for all $t \in [0, T_n]$. Claim 13 shows that σ_{+div} is a transverse section to f_n, which by the same reasoning, implies that $\phi_n(t, z_0)$ can never return to K_{+out} within the interval $[T_{ndiv}, T_n]$, and $\tilde{D}(z_n)$ is contained in $K_{+in} \cup \sigma_{+div} \cup \partial K_{+in} \cup \{z_+\}$ ($\subset K_{+in} \cup \partial K_{+div} \cup \partial K_{+in} \cup \{z_+\}$). Moreover, $l(z_+, z_n) \subset \sigma_{+div}$ is also a transverse section to f_n.

By Claim 2, the solutions of system Σ_n are unique, so that no two different paths can intersect [9, pp. 38]. Hence, no solution starting in $\tilde{D}(z_n) \setminus \tilde{\eta}_{\phi_n}(z_n)$ can reach $\tilde{\eta}_{\phi_n}(z_n)$, or exit $\tilde{D}(z_n)$ through the segment $\tilde{\eta}_{\phi_n}(z_n)$. This, together with the fact that $l(z_+, z_n)$ is a transverse section and $z_g \in l(z_+, z_n) \subset \tilde{D}(z_n) \setminus \tilde{\eta}_{\phi_n}(z_n)$, means that $\phi_n(t, z_g)$ can exit the region $\tilde{D}(z_n)$ only through the line segment $l(z_+, \phi_n(T_n, z_0)) \subset \partial K_{+in}$. By Fact 2, f_n and f_g coincide in $\tilde{D}(z_n) \subset K_{+in} \cup \partial K_{+div} \cup \partial K_{+in} \cup \{z_+\}$, so that $\phi_g(t, z_g) = \phi_n(t, z_g)$ at least until $\phi_n(t, z_g)$ exits $\tilde{D}(z_n)$, ie. until $t = T_g - T_{gdiv}$, where $\phi_g(T_g - T_{gdiv}, z_g) = \phi_g(T_g - T_{gdiv}, \phi_g(T_{gdiv}, z_0)) = \phi_g(T_g, z_0) \in \partial K_{+in}$. It follows that $\phi_g(t, z_g)$ can exit $\tilde{D}(z_n)$ only through the line segment $l(z_+, \phi_n(T_n, z_0))$, ie. $\phi_g(T_g, z_0) \in l(z_+, \phi_n(T_n, z_0))$, as desired. ■

Remark 15: In fact, a weaker version of Claim 15 (where the conclusion is that a $T_g \le T_n$ exists such that $\phi_g(T_g, z_0) \in l(z_+, \phi_n(T_n, z_0)) \cup \{\phi_n(T_n, z_0)\}$) suffices for the purpose of proving Proposition 3. The proof would have been shorter, as the condition $u_g(t) < u_n(t)$ for all $t \in (0, T]$ would be unnecessary. We present this marginally stronger result to confirm the intuitively reasonable conclusion. □

Consider a point $z_0 \in \partial K_{+in} \cap R_g$. From Fact 1, we have $\phi_g(t, z_0) \in \bar{K}$ for all $t \ge 0$. Recall the definition of σ_+ (see (10)), and let

$$t_{int} = \inf\{t \in (0, \infty) \mid \phi_g(t, z_0) \in \sigma_+\}.$$

In other words, t_{int} is the first time instant that the solution of the *GPAW compensated system* $\phi_g(t, z_0)$ intersects σ_+, or

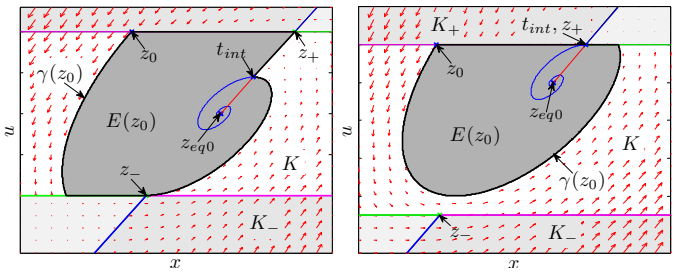


Fig. 4: Close path $\gamma(z_0)$ encloses region $E(z_0) \subset \bar{K}$. A case where the solution intersects ∂K_{-out}, then intersects σ_+ is shown on the left. A case where the solution intersects ∂K_{+out}, then intersects σ_+ at z_+ is shown on the right.

∞ if it does not intersect σ_+. If $t_{int} < \infty$, the path

$$\gamma_{int}(z_0) = \gamma_{int\phi_g}(z_0) \cup l(\phi_g(t_{int}, z_0), z_+) \cup \{z_+\} \cup l(z_0, z_+),$$

where

$$\gamma_{int\phi_g}(z_0) = \{z \in \mathbb{R}^2 \mid z = \phi_g(t, z_0), \forall t \in [0, t_{int}]\},$$

is well defined. Otherwise, the path

$$\gamma_0(z_0) = \gamma_{0\phi_g}(z_0) \cup \{z_{eq0}\} \cup \sigma_+ \cup l(z_0, z_+),$$

where

$$\gamma_{0\phi_g}(z_0) = \{z \in \mathbb{R}^2 \mid z = \phi_g(t, z_0), \forall t \geq 0\},$$

is well defined. Now, define the path $\gamma(z_0) \in \mathbb{R}^2$ by

$$\gamma(z_0) = \begin{cases} \gamma_{int}(z_0), & \text{if } t_{int} < \infty, \\ \gamma_0(z_0), & \text{otherwise,} \end{cases}$$

which can be verified to be closed and connected. Let the *open, bounded* region enclosed by $\gamma(z_0)$ be $E(z_0)$, and its closure be $\bar{E}(z_0)$. The region $E(z_0)$ is illustrated in Fig. 4.

Remark 16: Observe that $z_+ \in \sigma_+$. If $t_{int} < \infty$ and $\phi_g(t_{int}, z_0) = z_+ \in \sigma_+$, then $\gamma_{int}(z_0)$ reduces to

$$\gamma_{int}(z_0) = \gamma_{int\phi_g}(z_0) \cup l(z_0, z_+),$$

since $l(\phi_g(t_{int}, z_0), z_+) = l(z_+, z_+) = \{z_+\}$ and $z_+ \in \gamma_{int\phi_g}(z_0)$. □

The following result is analogous to Claims 7 and 8 combined, with respect to $\bar{E}(z_0)$.

Claim 16: If $z_0 \in \partial K_{+in} \cap R_g$, then $\bar{E}(z_0) \subset \bar{K}$ is a positive invariant set for system Σ_g. Moreover, $\bar{E}(z_0)$ is contained in the ROA of system Σ_g, and it must contain z_{eq0}, ie. $\bar{E}(z_0) \subset R_g$ and $z_{eq0} \in \bar{E}(z_0)$.

Proof: Let

$$\tilde{\sigma}_+ = \begin{cases} l(\phi_g(t_{int}, z_0), z_+) \cup \{z_+\}, & \text{if } t_{int} < \infty, \\ \sigma_+, & \text{otherwise.} \end{cases}$$

Observing from Fact 2 that f_n and f_g coincide on $\tilde{\sigma}_+ \subset K \cup \{z_+\}$, it can be verified as in the proof of Claim 7, that $\tilde{\sigma}_+$ is a transverse section to f_g and f_g always points into $\bar{E}(z_0)$ on $\tilde{\sigma}_+$. It is clear that $l(z_0, z_+) \subset \partial K_{+in}$ is also a transverse section to f_g, and that f_g always points into $\bar{E}(z_0)$ on $l(z_0, z_+)$. Both of these results show that any solution originating in $\bar{E}(z_0)$ cannot exit $\bar{E}(z_0)$ through the line segments $\tilde{\sigma}_+$ or $l(z_0, z_+)$. Furthermore, the solution $\phi_g(t, z_0)$ is unique due to Proposition 1, which implies that no solution originating in $\bar{E}(z_0)$ can exit it through the boundary $\gamma_{0\phi_g}(z_0)$ (or $\gamma_{int\phi_g}(z_0)$ as appropriate) (see Remark 4). These show that the region $\bar{E}(z_0)$ enclosed by $\gamma(z_0)$ must be a positive invariant set for system Σ_g. Fact 1 shows that $\phi_g(t, z_0) \in \bar{K}$ for all $t \geq 0$, which implies $\bar{E}(z_0) \subset \bar{K}$. This proves the first statement.

Since $z_0 \in R_g$, we have $\phi_g(t, z_0) \to z_{eq0}$ as $t \to \infty$. Since $\bar{E}(z_0)$ is a positive invariant set and $z_0 \in \bar{E}(z_0)$, we have $\phi_g(t, z_0) \in \bar{E}(z_0)$ for all $t \geq 0$. The conclusion $z_{eq0} \in \bar{E}(z_0)$ then follows from the fact that $\bar{E}(z_0)$ is *closed* and hence contains all its limit points.

It remains to show that $\bar{E}(z_0) \subset R_g$. If $\bar{E}(z_0) \cap \partial K_{+out} \neq \emptyset$, it can be verified that it must lie in the line segments $\gamma_{0\phi_g}(z_0)$ (or $\gamma_{int\phi_g}(z_0)$), ie. $(\bar{E}(z_0) \cap \partial K_{+out}) \subset \gamma_{0\phi_g}(z_0)$ (or $(\bar{E}(z_0) \cap \partial K_{+out}) \subset \gamma_{int\phi_g}(z_0)$). Hence any solution of Σ_g starting in $\bar{E}(z_0)$ that intersects ∂K_{+out} must intersect $\phi_g(t, z_0)$ at some time. Since $\lim_{t\to\infty} \phi_g(t, z_0) = z_{eq0}$, it follows from uniqueness of solutions that any solution starting from a point $\tilde{z} \in \bar{E}(z_0)$ that intersects ∂K_{+out} must converge to z_{eq0}, ie. $\tilde{z} \in R_g$. In similar manner, any solution starting from a point $\hat{z} \in \bar{E}(z_0)$ that intersects ∂K_{-out} must converge to z_{eq0}, ie. $\hat{z} \in R_g$. It suffices to consider solutions that do not intersect $\partial K_{+out} \cup \partial K_{-out}$, ie. solutions contained in $\tilde{E}(z_0) := \bar{E}(z_0) \setminus (\partial K_{+out} \cup \partial K_{-out})$.

It can be verified from Claim 6 that any equilibria of Σ_g apart from z_{eq0} contained in $\bar{E}(z_0)$ must lie in $\partial K_{+out} \cup \partial K_{-out}$. Then the only equilibrium point in $\tilde{E}(z_0)$ ($\subset \bar{E}(z_0)$) is z_{eq0}, which must be a stable node or focus. Observe that f_g is *continuously differentiable* in $\tilde{E}(z_0) \subset K \cup \partial K_{+in} \cup \partial K_{-in} \cup \{z_+, z_-\}$, so that Bendixson's Criterion [8, Lemma 2.2, pp. 67] applies in this region. As in the proof of Claim 8, Bendixson's Criterion [8, Lemma 2.2, pp. 67] and the absence of saddle points in $\tilde{E}(z_0)$ means that $\{z_{eq0}\}$ is the ω limit set of every solution contained in $\tilde{E}(z_0)$. Hence $\tilde{E}(z_0) \subset R_g$, and the conclusion follows. ■

The following is the main result of this subsection.

Proposition 3: The part of the ROA of the origin of system

Σ_n contained in $\mathbb{R}^2 \setminus \bar{K}$, is itself contained within the ROA of the origin of system Σ_g, ie. $(R_n \cap (\mathbb{R}^2 \setminus \bar{K})) \subset R_g$.

Proof: We need to show that if $z_0 \in R_n \cap (\mathbb{R}^2 \setminus \bar{K})$, then $z_0 \in R_g$. First, observe that $\mathbb{R}^2 \setminus \bar{K} = K_+ \cup K_-$, and $K_+ = K_{+out} \cup K_{+in} \cup \partial K_{+div}$. We will show that if $z_0 \in R_n \cap K_+$, then $z_0 \in R_g$. The proof where $z_0 \in R_n \cap K_-$ is similar. Let $z_0 \in R_n \cap K_+$. Since $z_0 \in R_n$ and $z_{eq0} \in K$, Fact 3 shows that $\phi_n(t, z_0)$ must intersect ∂K_{+in} at least once. Let T be the first time instant that $\phi_n(t, z_0)$ intersects ∂K_{+in}, so that $\phi_n(T, z_0) \in \partial K_{+in}$ and $\phi_n(t, z_0) \in K_+$ for all $t \in [0, T)$.

Consider when $z_0 \in R_n \cap (K_{+in} \cup \partial K_{+div}) \subset R_n \cap K_+$. We claim that $\phi_n(t, z_0)$ must be contained in $K_{+in} \cup \partial K_{+div}$ (and hence cannot enter K_{+out}) for all $t \in [0, T)$. Otherwise, $\phi_n(t, z_0)$ must intersect ∂K_{+div} at some finite time $\tilde{T} \in (0, T)$ and then pass into K_{+out}. Claims 13 and 14 shows that $\phi_n(t, z_0)$ must pass through ∂K_{+div} or σ_{+div}, which are transverse sections. By similar reasoning as in the proof of Claim 15, $\phi_n(t, z_0)$ can never return to K_{+in} during the interval $[\tilde{T}, T]$. In that case, $\phi_n(t, z_0)$ can never intersect ∂K_{+in}, which is a contradiction that establishes the immediate claim.

Since $\phi_n(t, z_0) \in K_{+in} \cup \partial K_{+div}$ for all $t \in [0, T)$, Fact 2 yields $\phi_g(t, z_0) = \phi_n(t, z_0)$ for all $t \in [0, T]$. Since $z_0 \in R_n$, we have $\phi_n(t, z_0) \in R_n$ for all $t \geq 0$. In particular, we have $\phi_n(T, z_0) = \phi_g(T, z_0) \in R_n \cap \partial K_{+in} \subset R_n \cap \bar{K}$. Proposition 2 then shows that $\phi_g(T, z_0) \in R_g$, so that $z_0 \in R_g$, as desired.

Next, consider when $z_0 \in R_n \cap K_{+out} \subset R_n \cap K_+$. Claim 15 shows that there exists a $T_g \in (0, T)$ such that $\phi_g(T_g, z_0) \in l(z_+, \phi_n(T, z_0)) \subset \partial K_{+in}$. Since $z_0 \in R_n$ and $\phi_n(T, z_0) \in \partial K_{+in}$, we have $\phi_n(T, z_0) \in R_n \cap \partial K_{+in} \subset R_n \cap \bar{K}$. Proposition 2 then shows that $\phi_n(T, z_0) \in R_g$. Observing that $\phi_n(T, z_0) \in \partial K_{+in} \cap R_g$, Claim 16 shows that $l(z_+, \phi_n(T, z_0)) \subset \bar{E}(\phi_n(T, z_0)) \subset R_g$. Then $\phi_g(T_g, z_0) \in l(z_+, \phi_n(T, z_0)) \subset R_g$ implies $z_0 \in R_g$, as desired.

Finally, by observing that

$$(R_n \cap (K_{+in} \cup \partial K_{+div})) \cup (R_n \cap K_{+out}) = R_n \cap K_+,$$

the conclusion follows. ■

C. Main Result

The following is the main result, which shows that the GPAW scheme can only maintain/enlarge the ROA of the uncompensated system. This establishes the GPAW scheme as a valid anti-windup method for this simple system.

Proposition 4: The ROA of the origin of system Σ_n is contained within the ROA of the origin of system Σ_g, ie. $R_n \subset R_g$.

Proof: Observing that

$$R_n = (R_n \cap \bar{K}) \cup (R_n \cap (\mathbb{R}^2 \setminus \bar{K})),$$

the result follows immediately from Propositions 2 and 3. ■

D. Numerical Examples

Here, we show numerical results on the *exact* ROAs of systems Σ_n and Σ_g. The reader is reminded that in these figures, the ROAs are to be interpreted as *open* sets, since ROAs must be *open* [8, Lemma 8.1, pp. 314]. Fig. 5a shows the case where $R_n = R_g$ for an open loop unstable system, together with two pairs of representative solutions, when the saturation constraints are symmetric, ie. $u_{max} = -u_{min}$. When the same system is subjected to *asymmetric saturation constraints*, the ROAs are illustrated in Fig. 5b. Clearly, the set containment $R_n \subset R_g$ is strict. In Fig. 5c, the ROAs are illustrated for an open loop strictly stable system with the nominal controller parameter chosen to satisfy $d \in (0, -a)$. Again, the set containment $R_n \subset R_g$ is strict.

Remark 17: Observe that the case of asymmetric saturation constraints arises whenever the objective is to regulate about an equilibrium not lying in $\{(x, u) \in \mathbb{R}^2 \mid u = 0\}$, and the system state is transformed such that the resulting equilibrium lies at the origin. □

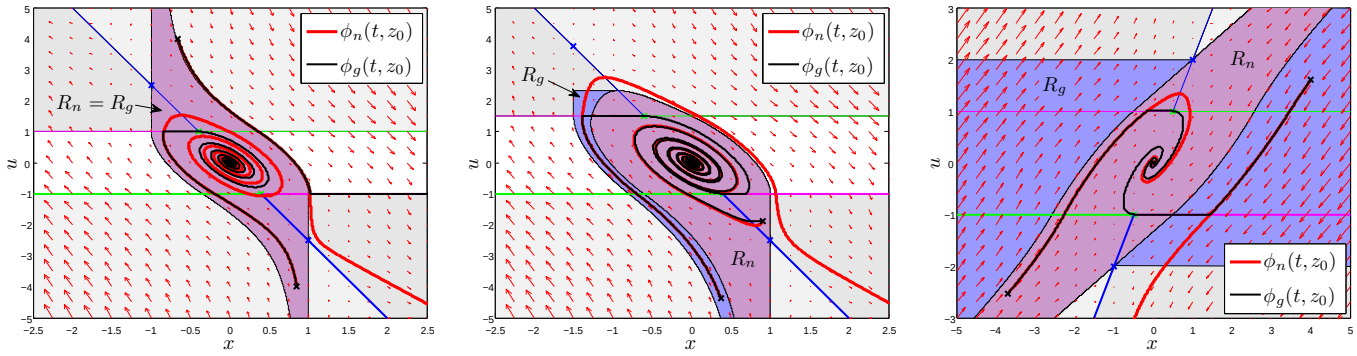
VII. A Paradigm Shift in Anti-windup Compensation

Here, we propose a new way of addressing the general anti-windup problem. To aid in the subsequent discussion, we present the next result, which states that the *nominal uncompensated system* achieves global asymptotic stability (GAS) and local exponential stability (LES) when both the open loop system and nominal controller are marginally or strictly stable.

Claim 17: If in addition to Assumption 1, both the open loop system and nominal controller are marginally or strictly stable ($a \leq 0$ and $d \leq 0$), then the origin of the nominal system Σ_n is globally asymptotically stable and locally exponentially stable.

Remark 18: This is the main reason why this case is not included in Fig. 5. □

Proof: The proof follows [30, Example 3.14, pp. 74 – 75] closely. First, the nominal system Σ_n is governed by the



(a) $a > 0, u_{max} = -u_{min}$ (b) $a > 0, u_{max} \neq -u_{min}$ (c) $a < 0, d \in (0, -a), u_{max} = -u_{min}$

Fig. 5: Numerical examples to illustrate the ROAs of systems Σ_n and Σ_g, which shows that the ROA containment $R_n \subset R_g$ of Proposition 4 can hold *strictly*. The vector field f_n is shown in the background, light purple regions represent R_n ($\subset R_g$), and light blue regions represent $R_g \setminus R_n$. In (a), the open loop system is unstable and the saturation limits are *symmetrical* ($a = 1$, $b = 1$, $c = -3$, $d = -1.2$, $u_{max} = -u_{min} = 1$), resulting in $R_n = R_g$. The pair of solutions starting at $z_0 = (0.85, -4) \in R_n \cap R_g$ converges to the origin, while the pair of solutions starting at $z_0 = (-0.66, 4) \notin R_n \cup R_g$ failed to converge to the origin. Cases (b) and (c) shows that $R_n \subset R_g$ holds *strictly*. Case (b) is identical with case (a), except with *asymmetric* saturation limits ($a = 1$, $b = 1$, $c = -3$, $d = -1.2$, $u_{max} = 1.5$, $u_{min} = -1$). Two pairs of solutions starting from $z_0 = (0.9, -1.9) \in R_n \cap R_g$ and $z_0 = (0.37, -4.37) \in R_g \setminus R_n$ are also included. A case where the open loop system is stable with an *unstable* controller is shown in (c) ($a = -1$, $b = 1$, $c = -1$, $d = 0.5$, $u_{max} = -u_{min} = 1$), together with two pairs of solutions starting from $z_0 = (-3.7, -2.54) \in R_n \cap R_g$ and $z_0 = (4, 1.6) \in R_g \setminus R_n$.

ordinary differential equations

$$\dot{x} = ax + b\,\mathrm{sat}(u),$$
$$\dot{u} = cx + du,$$

which can be rewritten as

$$\ddot{u} = (a + d)\dot{u} - adu + bc\,\mathrm{sat}(u). \tag{14}$$

Consider the continuously differentiable function

$$V(u, \dot{u}) = \tfrac{1}{2}\dot{u}^2 + \int_0^u ad\tau - bc\,\mathrm{sat}(\tau)\,d\tau = \tfrac{1}{2}\dot{u}^2 + \tilde{V}(u).$$

We will show that $V(u, \dot{u})$ is positive definite when $ad \geq 0$, which is implied by $a \leq 0$ and $d \leq 0$. Clearly, it is sufficient to show that $\tilde{V}(u) := \int_0^u ad\tau - bc\,\mathrm{sat}(\tau)\,d\tau$ is positive definite. When $u_{min} \leq u \leq u_{max}$, we have

$$\tilde{V}(u) = \int_0^u (ad - bc)\tau\,d\tau = \tfrac{1}{2}(ad - bc)u^2,$$

so that from (6), $\tilde{V}(u) > 0$ for all $u \in [u_{min}, u_{max}] \setminus \{0\}$. Next, consider when $u = \tilde{u} + u_{max} > u_{max}$, where $\tilde{u} > 0$. Direct computation yields

$$\begin{aligned}\tilde{V}(u) &= \tfrac{1}{2}adu^2 - \int_0^{u_{max}} bc\tau\,d\tau - \int_{u_{max}}^u bcu_{max}\,d\tau,\\ &= \tfrac{1}{2}adu^2 - \tfrac{1}{2}bcu_{max}^2 - bc\tilde{u}u_{max},\\ &= \tfrac{1}{2}ad(\tilde{u}^2 + 2\tilde{u}u_{max} + u_{max}^2) - \tfrac{1}{2}bcu_{max}^2 - bc\tilde{u}u_{max},\\ &= \tfrac{1}{2}ad\tilde{u}^2 + (ad - bc)\tilde{u}u_{max} + \tfrac{1}{2}(ad - bc)u_{max}^2.\end{aligned}$$

Clearly, when $ad \geq 0$, (6) implies $\tilde{V}(u) > 0$ for all $u > u_{max}$. The case when $u < u_{min}$ can be shown similarly. Hence $V(u, \dot{u})$ is positive definite. The above expressions also show that $V(u, \dot{u})$ is radially unbounded.

Taking the time derivative yields

$$\begin{aligned}\dot{V}(u, \dot{u}) &= \dot{u}\ddot{u} + (adu - bc\,\mathrm{sat}(u))\dot{u},\\ &= \dot{u}((a + d)\dot{u} - adu + bc\,\mathrm{sat}(u))\\ &\qquad + (adu - bc\,\mathrm{sat}(u))\dot{u},\\ &= (a + d)\dot{u}^2.\end{aligned}$$

By (5), we have $\dot{V}(u, \dot{u}) \leq 0$, ie. negative semidefinite.

To complete the proof for global asymptotic stability, it is sufficient to show that $\dot{V}(u, \dot{u}) \equiv 0$ implies $\dot{u} \equiv 0$ and $u \equiv 0$. The first condition is obtained immediately. When $\dot{u} \equiv 0$, (14) reduces to

$$\ddot{u} = -adu + bc\,\mathrm{sat}(u),$$

so that $\ddot{u}$ is nonzero as long as $u \not\equiv 0$. Hence only the trivial solution $u \equiv 0$, $\dot{u} \equiv 0$ can stay identically in the set $S = \{(u, \dot{u}) \in \mathbb{R}^2 \mid \dot{V}(u, \dot{u}) = 0\}$. By [8, Corollary 4.2, pp. 129], the origin of Σ_n is globally asymptotically stable. Local exponential stability of the origin follows immediately from Assumption 1. ■

Remark 19: Observe that (5) precludes $ad \geq 0$ being satisfied when either $a > 0$ or $d > 0$. □

Corollary 2: If in addition to Assumption 1, both the open

loop system and nominal controller are marginally or strictly stable ($a \leq 0$ and $d \leq 0$), then the origin of the GPAW compensated system Σ_g is globally asymptotically stable and locally exponentially stable.

Proof: Claim 17 shows that the origin z_{eq0} is globally asymptotically stable for system Σ_n, which implies $R_n = \mathbb{R}^2$. Proposition 4 then yields $R_g \supset R_n = \mathbb{R}^2$, which implies $R_g = \mathbb{R}^2$ and that the origin z_{eq0} is globally asymptotically stable for system Σ_g. Local exponential stability of the origin for system Σ_g follows immediately from Assumption 1. ■

Numerous results in the anti-windup literature are of the form of Corollary 2, ie. under some assumptions and applying some anti-windup method, some stability properties are achieved. Such results *sounds* impressive, and may indeed give some confidence in the application of the particular anti-windup method. However, we argue that it may not reveal any advantages of the anti-windup method. First, observe that for any meaningful anti-windup problem, local stability must be assumed to hold. Otherwise, the anti-windup problem is ill-posed. Any results asserting local stability are only restating the assumption. Observe from Claim 17 that for this example, the *uncompensated system* achieves GAS. While Corollary 2 asserts GAS, it tells nothing of any advantages gained by adopting the particular anti-windup method.

In contrast, the ROA containment result of Proposition 4 truly reflects an advantage of the GPAW scheme, namely, that the ROA of the system will always be maintained/enlarged by its application. As such, we propose this new paradigm to address the anti-windup problem, ie. results on the anti-windup compensated system *relative* to the uncompensated system.

Conclusion

We analyzed the gradient projection anti-windup (GPAW) scheme when applied to a constrained first order LTI system driven by a first order LTI controller, where the objective is to regulate the system state about the origin. Existence and uniqueness of solutions are assured using results from the projected dynamical systems literature, and equilibria are characterized. The main result of this report is that GPAW compensation applied to this simple system can only maintain/enlarge the system's region of attraction, which renders it a valid anti-windup method. The weaknesses of some qualitative results on anti-windup methods are illustrated, which motivated a new paradigm for addressing the anti-windup problem.

While these results are attractive, their applicability are severely limited. Extending these results to general MIMO nonlinear systems/controllers is a topic for future work.

Acknowledgments

The first author gratefully acknowledges Prof. Jean-Jacques Slotine for the numerous enlightening discussions that led to the development of the GPAW scheme. He also gratefully acknowledges the support of DSO National Laboratories, Singapore. Research funded in part by AFOSR grant FA9550-08-1-0086.

Appendix A
Derivation of GPAW Compensated Controller

Here, we derive the GPAW compensated controller (4) using the construction in [1], by enumerating all possibilities of the associated combinatorial optimization subproblem. The transformed nominal controller (3) is repeated here for ease of reference

$$\dot{u} = f_c(u, x) = cx + du.$$

First, there are two saturation constraints given by

$$h_1(u) := u - u_{max} \leq 0, \quad h_2(u) := -u + u_{min} \leq 0,$$

with associated *constant* gradient vectors

$$\nabla h_1 = \nabla h_1(u) = 1, \qquad \nabla h_2 = \nabla h_2(u) = -1.$$

There are three cases to consider, namely when $\mathcal{I}_{sat} = \{i \in \{1,2\} \mid h_i(u) \geq 0\} = \emptyset$, $\mathcal{I}_{sat} = \{1\}$, or $\mathcal{I}_{sat} = \{2\}$, corresponding to candidate solution sets $\mathcal{J} = \{\emptyset\}$, $\mathcal{J} = \{\emptyset, \{1\}\}$, or $\mathcal{J} = \{\emptyset, \{2\}\}$ respectively. Thus the possible candidate solutions to [1, subproblem (11)] are $\emptyset$, $\{1\}$, or $\{2\}$, and we have $N_\emptyset = [0]$, $N_{\{1\}} = [1]$, $N_{\{2\}} = [-1]$ in accordance to the definition in [1].

Remark 20: Clearly, the case $\mathcal{I}_{sat} = \{1, 2\}$ can never occur due to $u_{min} < u_{max}$. □

Choosing any scalar $\Gamma > 0$ and defining

$$f_{\mathcal{I}}(u, x) = R_{\mathcal{I}}(u) f_c(u, x) = R_{\mathcal{I}}(cx + du),$$

where

$$R_{\mathcal{I}} = \begin{cases} I - \Gamma N_{\mathcal{I}} \left(N_{\mathcal{I}}^{\mathrm{T}} \Gamma N_{\mathcal{I}} \right)^{-1} N_{\mathcal{I}}^{\mathrm{T}}, & \text{if } |\mathcal{I}| > 0, \\ I, & \text{otherwise,} \end{cases}$$

the above can be evaluated for any $\mathcal{I} \in \{\emptyset, \{1\}, \{2\}\}$ to be

$$f_{\mathcal{I}}(u, x) = \begin{cases} cx + du, & \text{if } \mathcal{I} = \emptyset, \\ 0, & \text{if } \mathcal{I} = \{1\} \text{ or } \mathcal{I} = \{2\}. \end{cases}$$

The GPAW compensated controller is then given by

$$\dot{u} = f_{\mathcal{I}^*}(u, x) = \begin{cases} cx + du, & \text{if } \mathcal{I}^* = \emptyset, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathcal{I}^*$ is the solution of [1, subproblem (11)], translated as the equivalent combinatorial optimization problem

$$\begin{aligned} &\max_{\mathcal{I}\in\mathcal{J}} J(\mathcal{I}) := (cx+du) f_{\mathcal{I}}(u,x), \\ &\text{subject to } N^{\mathrm{T}}_{\mathcal{I}_{sat}\setminus\mathcal{I}} f_{\mathcal{I}}(u,x) \leq 0. \end{aligned} \tag{15}$$

Observe that the GPAW compensated controller is *independent* of Γ, and the rank (or linear independence) condition in [1, subproblem (11)] is automatically satisfied for this simple system.

Now, consider the three cases mentioned previously. Clearly, when $\mathcal{I}_{sat} = \emptyset$, ie. when no constraints are violated, $\mathcal{I}^* = \emptyset$ is the optimal solution to problem (15). When $\mathcal{I}_{sat} = \{1\}$, we have $\mathcal{J} = \{\emptyset, \{1\}\}$ and the objective function evaluates to

$$J(\mathcal{I}) = \begin{cases} (cx+du)^2, & \text{if } \mathcal{I} = \emptyset, \\ 0, & \text{if } \mathcal{I} = \{1\}. \end{cases}$$

When $cx + du = 0$, it is immaterial whether $\mathcal{I}^*$ is chosen as $\emptyset$ or $\{1\}$. When $cx + du \neq 0$, inspection of the preceding and (15) shows that $\mathcal{I}^* = \{1\}$ only when $N^{\mathrm{T}}_{\mathcal{I}_{sat}\setminus\emptyset} f_\emptyset(u,x) = N^{\mathrm{T}}_{\mathcal{I}_{sat}} f_\emptyset(u,x) = N^{\mathrm{T}}_{\{1\}} f_\emptyset(u,x) = cx + du > 0$. By similar arguments, we can show that when $\mathcal{I}_{sat} = \{2\}$, the solution to problem (15) is $\mathcal{I}^* = \{2\}$ only when $cx+du = 0$, or $cx+du \neq 0$ and $N^{\mathrm{T}}_{\mathcal{I}_{sat}\setminus\emptyset} f_\emptyset(u,x) = N^{\mathrm{T}}_{\mathcal{I}_{sat}} f_\emptyset(u,x) = N^{\mathrm{T}}_{\{2\}} f_\emptyset(u,x) = -(cx+du) > 0$. Collecting these individual cases, we can write the GPAW compensated controller as

$$\dot{u} = \begin{cases} 0, & \text{if } \mathcal{I}_{sat} = \{1\} \text{ and } cx+du > 0, \\ 0, & \text{if } \mathcal{I}_{sat} = \{2\} \text{ and } cx+du < 0, \\ cx+du, & \text{otherwise,} \end{cases}$$

which is equivalent to (4).

Appendix B
Simplifying Some Logical Statements

In some of the proofs in this report, we need to assert the truth of statements of the form "if $z \in \alpha$ and $z \in \beta$, then $z \in \gamma$". Here, we show explicitly that this statement is equivalent to "if $z \in \alpha \setminus \gamma$, then $z \notin \beta$". Note that $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$, represents logical *negation* ("NOT" operator), *conjunction* ("AND" operator), *disjunction* ("OR" operator), *implication*, and *equivalence* respectively.

Let

$$A = (z \in \alpha), \qquad B = (z \in \beta), \qquad C = (z \in \gamma),$$

so that the original statement is equivalent to $(A \wedge B) \Rightarrow C$. We can use the equivalence $(\phi \Rightarrow \theta) \Leftrightarrow (\neg\phi \vee \theta)$ [31, Figure 7.11, pp. 210] to write the original statement as

$$\begin{aligned} (A \wedge B) \Rightarrow C &\Leftrightarrow \neg(A \wedge B) \vee C \Leftrightarrow \neg A \vee \neg B \vee C, \\ &\Leftrightarrow \neg A \vee C \vee \neg B \Leftrightarrow \neg(A \wedge \neg C) \vee \neg B, \\ &\Leftrightarrow (A \wedge \neg C) \Rightarrow \neg B. \end{aligned}$$

In other words, the original statement is equivalent to "if $z \in \alpha$ and $z \notin \gamma$, then $z \notin \beta$", or more compactly, "if $z \in \alpha \setminus \gamma$, then $z \notin \beta$".

Moreover, observe that we can always replace A by more complex statements to get an analogous equivalence relation. For example, if $A = (D \vee E) \wedge F$, then

$$((D \vee E) \wedge F \wedge B \Rightarrow C) \Leftrightarrow ((D \vee E) \wedge F \wedge \neg C \Rightarrow \neg B).$$

In fact, the more complex form is often used in this report.

Appendix C
A Variant of the Comparison Lemma

Here, we present a variant of the Comparison Lemma [8, Lemma 3.4, pp. 102 – 103], where the conclusion results in a *strict* inequality. It is a direct consequence of uniqueness of solutions of the *scalar* differential equation, with an application of the original Comparison Lemma.

Lemma 1 (Strict Comparison Lemma): Consider the scalar differential equation

$$\dot{u} = f(t,u), \qquad u(t_0) = u_0, \tag{16}$$

where $f(t,u)$ is continuous in t and locally Lipschitz in u, for all $t \geq 0$ and all $u \in J \subset \mathbb{R}$, J a *connected* set. Let $[t_0, T)$ (T could be infinity) be the maximal interval of existence of the solution $u(t)$, and suppose $u(t) \in J$ for all $t \in [t_0, T)$. Let $v(t)$ be a continuous function whose upper right-hand derivative $D^+v(t)$ satisfies the differential inequality

$$D^+v(t) \leq f(t, v(t)), \qquad v(t_0) < u_0,$$

with $v(t) \in J$ for all $t \in [t_0, T)$. Then $v(t) < u(t)$ for all $t \in [t_0, T)$.

Remark 21: Observe that the fundamental qualitative difference with [8, Lemma 3.4, pp. 102 – 103] is the *strict* inequality of the initial condition $v(t_0) < u_0$, and the conclusion $v(t) < u(t)$ for all $t \in [t_0, T)$. The requirement of J being a *connected* set is purely technical, as seen in the proof. □

Proof: Consider the initial value problem

$$\dot{w} = f(t,w), \qquad w(t_0) = v(t_0) < u_0. \tag{17}$$

With the assumptions, [8, Theorem 3.1, pp. 88 – 89] implies existence and uniqueness of solutions of (16) and (17). Let

$[t_0, T_w)$ be the maximal interval of existence of the solution $w(t)$ such that $w(t) \in J$ for all $t \in [t_0, T_w)$. Define $\tilde{T} := \min\{T, T_w\}$.

We claim that $w(t) \neq u(t)$ for all $t \in [t_0, \tilde{T})$ (due to $w(t_0) \neq u(t_0)$). Otherwise, there exists a $\hat{T} \in [t_0, \tilde{T})$ such that $w(\hat{T}) = u(\hat{T})$. By solving (16) and (17) *backwards in time* from $t = \hat{T}$ to $t = t_0$, we obtain $w(t_0) = u(t_0)$ due to uniqueness of solutions. This contradicts $w(t_0) \neq u(t_0)$ and establishes the claim.

Since $w(t_0) < u(t_0)$, and $w(t) \neq u(t)$ for all $t \in [t_0, \tilde{T})$, continuity of both $w(t)$ and $u(t)$ shows that $w(t) < u(t)$ holds with *strict* inequality for all $t \in [t_0, \tilde{T})$. The Comparison Lemma [8, Lemma 3.4, pp. 102 – 103] applied to (17) and the differential inequality yields $v(t) \leq w(t)$ for all $t \in [t_0, \tilde{T})$. Then we have $v(t) \leq w(t) < u(t)$ for all $t \in [t_0, \tilde{T})$. This, together with the connectivity of J and the condition $u(t), v(t) \in J$ for all $t \in [t_0, T)$ implies $w(t) \in J$ for all $t \in [t_0, T)$, ie. $T_w \geq T$ and $\tilde{T} = T$. Hence the conclusion $v(t) < u(t)$ holds with *strict* inequality for all $t \in [t_0, T)$. ■

REFERENCES

[1] J. Teo and J. P. How, "Anti-windup compensation for nonlinear systems via gradient projection: Application to adaptive control," in *Proc. 48th IEEE Conf. Decision and Control & 28th Chinese Control Conf.*, Shanghai, China, Dec. 2009, pp. 6910 – 6916.

[2] S. Tarbouriech and M. Turner, "Anti-windup design: an overview of some recent advances and open problems," *IET Control Theory Appl.*, vol. 3, no. 1, pp. 1 – 19, Jan. 2009.

[3] F. Morabito, A. R. Teel, and L. Zaccarian, "Nonlinear antiwindup applied to Euler-Lagrange systems," *IEEE Trans. Robot. Autom.*, vol. 20, no. 3, pp. 526 – 537, Jun. 2004.

[4] J. B. Rosen, "The gradient projection method for nonlinear programming. part I. linear constraints," *J. Soc. Ind. Appl. Math.*, vol. 8, no. 1, pp. 181 – 217, Mar. 1960.

[5] ——, "The gradient projection method for nonlinear programming. part II. nonlinear constraints," *J. Soc. Ind. Appl. Math.*, vol. 9, no. 4, pp. 514 – 532, Dec. 1961.

[6] K. J. Åström and L. Rundqwist, "Integrator windup and how to avoid it," in *Proc. American Control Conf.*, Pittsburgh, PA, Jun. 1989, pp. 1693 – 1698.

[7] S. Sastry, *Nonlinear Systems: Analysis, Stability, and Control*, ser. Interdiscip. Appl. Math. New York, NY: Springer, 1999, vol. 10.

[8] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[9] J. K. Hale, *Ordinary Differential Equations*, 2nd ed. Mineola, NY: Dover, 1997.

[10] M. Vidyasagar, *Nonlinear Systems Analysis*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1993.

[11] O. Hájek, *Control Theory in the Plane*, 2nd ed., ser. Lect. Notes Control Inf. Sci. Berlin, Germany: Springer, 2009, vol. 153.

[12] R. Mantri, A. Saberi, and V. Venkatasubramanian, "Stability analysis of continuous time planar systems with state saturation nonlinearity," *IEEE Trans. Circuits Syst. I*, vol. 45, no. 9, pp. 989 – 993, Sep. 1998.

[13] J. Alvarez, R. Suárez, and J. Alvarez, "Planar linear systems with single saturated feedback," *Syst. Control Lett.*, vol. 20, no. 4, pp. 319 – 326, Apr. 1993.

[14] J.-Y. Favez, P. Mullhaupt, B. Srinivasan, and D. Bonvin, "Attraction region of planar linear systems with one unstable pole and saturated feedback," *J. Dyn. Control Syst.*, vol. 12, no. 3, pp. 331 – 355, Jul. 2006.

[15] T. Hu, L. Qiu, and Z. Lin, "Stabilization of LTI systems with planar anti-stable dynamics using saturated linear feedback," in *Proc. 37th IEEE Conf. Decision and Control*, Tampa, FL, Dec. 1998, pp. 389 – 394.

[16] T. Hu, Z. Lin, and L. Qiu, "Stabilization of exponentially unstable linear systems with saturating actuators," *IEEE Trans. Autom. Control*, vol. 46, no. 6, pp. 973 – 979, Jun. 2001.

[17] M. L. Corradini, A. Cristofaro, and F. Giannoni, "Sharp estimates on the region of attraction of planar linear systems with bounded controls," in *Proc. 48th IEEE Conf. Decision and Control & 28th Chinese Control Conf.*, Shanghai, China, Dec. 2009, pp. 5345 – 5350.

[18] F. Liu, G. T.-C. Chiu, E. S. Hamby, and Y. Eun, "Time maximum control for a class of single-input planar affine control systems and constraints," in *Proc. 48th IEEE Conf. Decision and Control & 28th Chinese Control Conf.*, Shanghai, China, Dec. 2009, pp. 5045 – 5050.

[19] P. Dupuis and A. Nagurney, "Dynamical systems and variational inequalities," *Ann. Oper. Res.*, vol. 44, no. 1, pp. 7 – 42, Feb. 1993.

[20] D. Zhang and A. Nagurney, "On the stability of projected dynamical systems," *J. Optim. Theory Appl.*, vol. 85, no. 1, pp. 97 – 124, Apr. 1995.

[21] A. Nagurney and D. Zhang, *Projected Dynamical Systems and Variational Inequalities with Applications*, ser. Int. Ser. Oper. Res. Manag. Sci. Norwell, MA: Kluwer, 1996.

[22] M.-G. Cojocaru and L. B. Jonker, "Existence of solutions to projected differential equations in Hilbert spaces," *Proc. Amer. Math. Soc.*, vol. 132, no. 1, pp. 183 – 193, Jan. 2004.

[23] B. Brogliato, A. Daniilidis, C. Lemaréchal, and V. Acary, "On the equivalence between complementarity systems, projected systems and differential inclusions," *Syst. Control Lett.*, vol. 55, no. 1, pp. 45 – 51, Jan. 2006.

[24] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*. West Sussex, England: Wiley, 1996.

[25] J. Teo and J. P. How, "Geometric properties of gradient projection anti-windup compensated systems," in *Proc. American Control Conf.*, Baltimore, MD, Jun./Jul. 2010, to appear.

[26] A. S. Hodel and C. E. Hall, "Variable-structure PID control to prevent integrator windup," *IEEE Trans. Ind. Electron.*, vol. 48, no. 2, pp. 442 – 451, Apr. 2001.

[27] A. F. Filippov, *Differential Equations with Discontinuous Righthand Sides*, ser. Math. Appl. Dordrecht, Netherlands: Kluwer, 1988.

[28] A. A. Andronov, A. A. Vitt, and S. E. Khaikin, *Theory of Oscillators*, ser. Int. Ser. Monogr. Phys. Oxford, England: Pergamon Press, 1966, vol. 4.

[29] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, ser. Appl. Math. Sci. New York, NY: Springer, 2002, vol. 42.

[30] J.-J. Slotine and W. Li, *Applied Nonlinear Control*. Upper Saddle River, NJ: Prentice Hall, 1991.

[31] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed., ser. Artif. Intell. Upper Saddle River, NJ: Prentice Hall, 2003.