Inverse Optimal Safety Filters

Miroslav Krstic, *Fellow, IEEE*

***Abstract*—Control barrier function quadratic programs (QP) safety filters are pointwise minimizers of the control effort at a given state, i.e., myopically optimal at each time. But are they optimal over the entire infinite time horizon? What does it mean for a controlled system to be "optimally safe" as opposed to, conventionally "optimally stable?" When disturbances, deterministic and stochastic, have unknown upper bounds, how should safety be defined to allow a graceful degradation under disturbances? Can safety filters be designed to guarantee such weaker safety properties as well as the optimality of safety over the infinite time horizon? We pose and answer these questions for general systems affine in control and disturbances and illustrate the answers using several examples. In the process, using the existing QP safety filters, as well as more general safety-ensuring feedbacks, we generate entire families of safety filters that are optimal over the infinite horizon although they are conservative (favoring safety over "liveness") relative to the standard QP.**

***Index Terms*—Control barrier functions, inverse optimality, safety filters.**

I. Introduction

A. Control Barrier Functions (CBFs): A Few Highlights

IT WAS in two ways that the 2014 paper [4], along with its later journal version [5], marked a watershed in the study of nonlinear control systems under state constraints.

First, by advancing the notion of control barrier functions (CBF) proposed in [50], it laid the foundation for a Lyapunov-like alternative to constraint-handling control design methodssuch as classical optimal control, model predictive control (MPC) [40], or barrier Lyapunov functions (BLFs) [48]. While similar in name to CBFs, BLFs represent a more conservative approach in which the system is actively repelled from the boundary, as opposed to being just slowed down in its approach to the boundary. Furthermore, neither MPC nor BLFs entail the notion of a nominal control, as the primary purpose for the application of the input, from which the constraint-handling design should deviate only when a safety violation is imminent. Second, the authors in [4] and [5] proposed, following inspiration from [17], that the conflict between safety and the said nominal control (equilibrium stabilization, trajectory tracking, or mere open-loop forcing of the system) be "mediated" using a quadratic program (QP), in which the deviation of the actual control input from the nominal input is penalized quadratically, whereas the linear inequality constraint comes from the linearity in control of the nonnegative sum of the derivative of the CBF with an appropriate decay margin that limits the rate of approach to the barrier. This approach to imparting safety on a controlled system, while also obeying the system operator's intent, has been the most influential legacy of [4] and [5]. Virtually all the work on CBF-based safety maintenance employs some form of QP-based redesigns of the nominal control, often referred to as "safety filters."

CBFs have since been used in a range of domains, including multiagent robotics [18], [43], [49], automotive systems [4], [39], [55], robust safety [23], [25], [53], delay systems [1], [22], [34], [37], and stochastic systems [13], [38], [44].

Since CBFs define constraints and, as such, represent system outputs, when paired with system inputs they have relative degrees. For example, a position constraint, such as a relative distance between cars on the road, is of relative degree two in reference to an idealized accelerator input on a car but of relative degree three or higher in reference to the actual engine throttle input. CBFs of high relative degree, under that name, were first studied in the 2015 articles [20], [51] with progress following in [10], [35], [54], [52], and continuing. However, control designs for specific CBF of arbitrarily high relative degree already appear in the 2006 article [28], which presents backstepping designs for the regulation to the boundary of the safe set, referred to, at that time, as "nonovershooting control."

B. L_gh "Safety Filters"

QP-based safety filters are reminiscent of the 1980's-era parameter projection used in adaptive control [29, Appendix E], which defines the safe set through a "zeroing CBF." Between parameter projection and QP-based safety filters, there are two differences and one key similarity. One difference is that, in parameter estimation, the plant is simply a vector integrator (of the update law), as opposed to being a general nonlinear system affine in control. The other difference is that parameter projection is an extreme (discontinuous) form of a QP-based safety filter: projection lets the nominal update proceed unaltered up to the boundary of the safe set and then tangentially projects the update, allowing the trajectory to slide along the boundary if the nominal update directs the estimates outward. As for the key similarity between parameter projection and QP safety filters, projection also employs a CBF, as well as a QP. As a result, it has an L_gh factor, a hallmark of CBF-QP. More on this in Section X.

A factor of L_gh is a tell-tale sign of potential optimality—not mere pointwise optimality, at a given point x in the state space,

Manuscript received 19 December 2022; accepted 13 May 2023. Date of publication 22 May 2023; date of current version 29 December 2023. This work was supported in part by NSF under Grant ECCS-2151525, in part by AFOSR under Grant FA9550-22-1-0265, and in part by ONR under Grant N00014-23-1-2376. Recommended by Associate Editor T. Faulwasser.

The author is with the Department of Mechanical and Aerospace Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: krstic@ucsd.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2023.3278788.

Digital Object Identifier 10.1109/TAC.2023.3278788

but optimality over the infinite time horizon. The so-called "L_gV controllers" have a storied history in nonlinear stabilization. Sontag [47] "universal formula" was the first generally applicable L_gV controller and is both pointwise and infinite-horizon optimal. Sepulchre et al. [45] produced a collection of results with such "damping controllers" and showed that every L_gV controller—not just Sontag's formula—is optimal with respect to a meaningful cost functional if multiplied by a factor of two or more, which, in particular, indicates the controller's infinite gain margin. They also proved a nonlinear version of a 60° phase margin: an L_gV controller remains stabilizing when applied through any dynamical system of the form $\mathrm{Id}+\mathcal{P}$, where Id denotes identity and $\mathcal{P}$ denotes any strictly passive nonlinear system, which need not be input-to-state stable.

Such properties of L_gV controllers inspired their further development under uncertainties. In [27], for systems affine in control and disturbances, inverse optimal controllers were designed that solve a zero-sum game problem, in which the disturbance maximizes and the control minimizes a meaningful cost. In [21] and [36], global inverse optimality was augmented with local direct optimality. In [15] and [16], stochastic inverse optimal designs were introduced: L_gV controllers for inverse optimal stabilization in probability in [16] and controllers that are inverse optimal for a zero-sum game relative to the unknown covariance acting as the opposing player in [15]. Finally, in [31], *adaptive* L_gV controllers were designed that minimize a penalty not only on the plant's state and the input, but also on the parameter estimation error—thus far the only pairings of controllers and parameter estimators that are not merely optimal "asymptotically" but over the entire time horizon. In each of [15], [16], [27], and [31], L_gV controllers are designed not only for some classes of systems but for all suitably stabilizable systems, using Sontag-type formulae.

Given the L_gh form of the CBF-QP safety filters [4], [5], it is imperative to ask the following questions. Are the CBF-QP safety filters inverse optimal? If not, can they be made optimal with respect to some meaningful cost functionals? What is meaningful to penalize when "mediating" safety and the execution of the user's nominal control design?

To answer these questions, let us consult intuition. First, let us note that CLFs and CBFs are not the opposites of each other: CLF is an energy-like, or norm-like function, whereas a CBF is a system output. However, for both CLFs and CBFs, we are interested in their decays and growths. While the decay of a CLF indicates convergence to an equilibrium, i.e., an improvement in desired performance, the increase of a CBF indicates movement away from the dangerous boundary of the safe set, i.e., an improvement in safety. Hence, optimization should reward an increase in safety. Another hint comes from terminology: if the L_gV controllers got nicknamed the "damping controllers" because they enhance the negativity of $\dot{V}$, the CBF-QP safety filters, which *reduce* the negativity of $\dot{h}$, should be nicknamed "antidampers" among safety filters. In fact, the CBF-QP safety filters act precisely as pointwise worst-case disturbances, not unlike the optimal disturbances in H_∞ control.

In summary, optimal safety filters should be maximizing a reward function that is 1) proportional to the CBF and 2) negative definite in the deviation between the control applied and the nominal control. In plain language, optimality should reward both safety and close adherence to the nominal control.

Let us now return to the question—is CBF-QP inverse optimal? It is not. It is only optimal in a myopic sense, pointwise in x, but not over the infinite horizon. Infinite-horizon optimality has been pursued in [3], [2], [11], and [14] but toward achieving optimal stabilization, not *optimal safety*.

Can we design safety filters that have a property of inverse optimality? Yes and, in the absence of a disturbance, such a redesign amounts to little more than multiplying the QP modification to the nominal control by a factor of two or higher. Plainly speaking, doubling the antidamping of the CBF, i.e., doubling safety, imparts inverse optimality.

C. Nonlinear Systems With Disturbances: Deterministic and Stochastic

Under deterministic disturbances, two main ideas have emerged. Robust CBFs [23] ensure safety under a disturbance with a known bound. In input-to-state safety (ISSf) [25], which mirrors input-to-state stability [46], the disturbance is bounded but potentially arbitrarily large and, being also unvanishing, may take the system outside of the safe set. Hence, the CBF h may assume negative values but in proportion to the size of the disturbance, with a class $\mathcal{K}$ gain from the disturbance to the "safety violation" $-h$.

Controllers that render the safety violation $-h$ proportional to the disturbance are introduced in the 2006 work on nonovershooting control [28] with a backstepping design for a high relative degree CBF.

In the stochastic case, a general CBF-based safety analysis is presented in [13]. A mean-nonovershooting tracking design for stochastic strict-feedback systems is given in [30].

In this article, we tackle four questions related to systems with disturbances: Two of the questions are the designs of QP-based safety filters for general nonlinear systems affine in deterministic or stochastic disturbances. The other two questions are the inverse optimal versions of safety filters under stochastic and deterministic disturbances.

But what does inverse optimality mean in the presence of disturbances? Disturbances not subject to a known upper bound dictate that optimality take a form of a two-player game, between the safety filter and the disturbance. While the safety filter's goal is to maintain safety, the goal of the disturbance is to erode safety, while investing as little of its energy as possible. This leads to cost functionals positive definite in the disturbance and proportional to the CBF, with the goal of the disturbance to minimize such a cost.

Contribution Summary: The expansion of the "safety filter toolkit," which we offer here, relative to the introductory CBF-QP [5], is displayed in Table I. We design safety filters that are deterministic and stochastic disturbance-based versions of the CBF-QP design. We also provide their modifications that ensure inverse optimality. Our safety filters are Nash equilibrium strategies, in balance with the Nash equilibrium strategies of the disturbances.

D. Safety Framework

We are unconcerned with stability in this article. We consider a hierarchical scenario comprising the following:

1) at the bottom layer, an "operator" $\mathcal{O}$, who only commands setpoints or open-loop reference signals;

TABLE I
DESIGNS THAT EXPAND THE "SAFETY FILTER TOOLKIT" RELATIVE TO THE INTRODUCTORY CBF-QP [5]

	deterministic		stochastic	
	no disturbance	with disturbance	known noise covariance	unknown noise covariance
QP and *not* inv. opt.	[5]	Thm 2	Thm 6	Thm 9
QP and inv. opt.	Cor 1 and 2	Thm 5	Thm 8	
non-QP and inv. opt.		Thm 4	Thm 7	Thm 10

2) at the middle layer, a designer $\mathcal{N}$ of a nominal feedback law u_0, which fulfills $\mathcal{O}$'s command in the absence of state constraints;
3) at the top layer, a designer $\mathcal{S}$ of a safety filter $\bar{u}$, which ensures safety for a given nominal u_0.

The barrier function $h(x)$ is known only to $\mathcal{S}$. The scenario considered in CBF-CLF-QP [4], [23] has the setpoint in the safe set. We allow $\mathcal{O}$ to possibly command operation outside of the safe set (unknown to $\mathcal{O}$) or unstable operation (Example 2), and this makes the stability issue moot.

E. Preview of Inverse Optimal Safety Filters: A Scalar Example

Consider the scalar system

$$\dot{x} = u\,, \quad h(x) = -x\,. \tag{1}$$

Merely maintaining the positive invariance of $\{x < 0\}$ is achievable with trivial $u = 0$ and even with destabilizing $u = x$. A good safety filter should keep u close to the nominal u_0 when $-x > 0$ is comparatively large, i.e., when x is far from the boundary $x = 0$.

The QP solution $\bar{u}_{\rm QP} = \min\{0, -u_0 - x\}$ gives the safety filter $u = u_{\rm QP} = u_0 + \bar{u}_{\rm QP} = \min\{u_0, -x\}$, which ensures safety with $\dot{h} \geq -h$ and is pointwise optimal in x but is not optimal over the interval $0 \leq t < \infty$. However, the modified QP safety filter $u = u^*_{\rm QP} = u_0 + \mathbf{2}\bar{u}_{\rm QP} = u_0 + 2\min\{0, -u_0 - x\} = \min\{u_0, -u_0 - 2x\}$, namely

$$u = u^*_{\rm QP} = -x - |u_0 + x| \tag{2}$$

not only ensures safety but also, as we shall see in the general results later, *maximizes* the cost functional $-x(+\infty) + \int_0^\infty \left(\min\{-x, u_0\} - \frac{(u-u_0)^2}{4\max\{0, u_0+x\}}\right) {\rm d}t$, and, equivalently, *minimizes* the cost functional

$$x(+\infty) + \int_0^\infty \left(\max\{x, -u_0\} + \frac{(u-u_0)^2}{4\max\{0, u_0+x\}}\right) {\rm d}t \tag{3}$$

where we have simply suppressed the dependence on t in $x(t), u(t), u_0(x(t), t)$ under the integrals for the sake of clarity.

The functional (3) is meaningful. The term $x(+\infty)$ is a "terminal safety violation" cost, and the term $\max\{x, -u_0\}$ under the integral is a running safety violation cost. The term $(u-u_0)^2$ penalizes the deviation of u from u_0, and its denominator inflicts an infinite penalty on u for possibly not remaining at exactly the nominal u_0 when $u_0 < -x$, which is when the nominal control is acting on its own to push the state away from the boundary $x = 0$. The value function of (3) is the "safety violation" $+x$, which means that the optimizing safety filter results in the optimal cost $+x_0 < 0$. In summary, (3) incentivizes both safety and "liveness."

If the reader is unsettled by the nonsmoothness of $\max$ in (3), or simply put off by the dogmatism of QP/min-norm control, an alternative inverse optimal safety filter is the Sontag formula-inspired $\bar{u}_{\rm S} = -(u_0 + x + \sqrt{(u_0+x)^2+1})$, which gives $u = u^*_{\rm S} = u_0 + \bar{u}_{\rm S}$, namely

$$u = u^*_{\rm S} = -x - \sqrt{(u_0+x)^2+1} \tag{4}$$

which *minimizes*

$$x(+\infty) + \int_0^\infty \left[\frac{-u_0 + x + \sqrt{(u_0+x)^2+1}}{2} + \frac{1}{2}\frac{(u-u_0)^2}{u_0 + x + \sqrt{(u_0+x)^2+1}}\right] {\rm d}t \tag{5}$$

and hence, such as (3), also maximizes safety and minimizes $u - u_0$. Note the similarity between the optimal QP filter (2) and the slightly more conservative optimal Sontag filter (4). Another variation on the Sontag formula is the safety filter $u = u_{\rm S} = u_0 + \frac{1}{2}\bar{u}_{\rm S} = \frac{1}{2}(u_0 - x - \sqrt{(u_0+x)^2+1})$, which is not inverse optimal but guarantees safety with $\dot{h} \geq -h$.

Fig. 1 illustrates the following properties that accompany the safety filters in this article:

1) the nominal input is endowed with optimality away from the boundary;
2) the cost (safety deficit) increases near the boundary;
3) the nominal input is overriden near the boundary when the input is safety-reducing.

F. Contributions and Organization

The rest of this article is organized as follows. Sections II–V deal with deterministic disturbances. The main result on inverse optimality for safety filters is in Section V. Safety filters of the special QP form, under disturbances, are presented in Section IV and their inverse optimal versions at the end of Section V.

For the reader less interested in the effects of disturbances (or overwhelmed by them), Section VI specializes the inverse optimality results to the disturbance-free case. It is in this section that the points of this article are most transparently evident. The reciprocal CBF (RCBF) formulation of inverse optimal safety filter design in this section is the closest to the traditional notion of optimal control—the task of control is minimization.

Stochastic systems are dealt with in Sections VII–IX. Stochastic safety filters, stochastic QP formulae, and inverse optimal achievement of safety under stochastic disturbances are all new notions in the literature. In addition, the notion of stochastic safety under nonunity covariance, where covariance is time dependent and of unknown bound, which is the subject of Section IX, is a new topic in the safety literature. Sections VIII and IX are written in a contrasting fashion: the former dealing with the easier but still novel case of inverse optimal safety filters for unity-intensity stochastic disturbances, and the latter dealing with the same topics but with covariance whose intensity is arbitrary and incorporated in the cost functional—rewarded for making the system less safe but penalized when its energy is

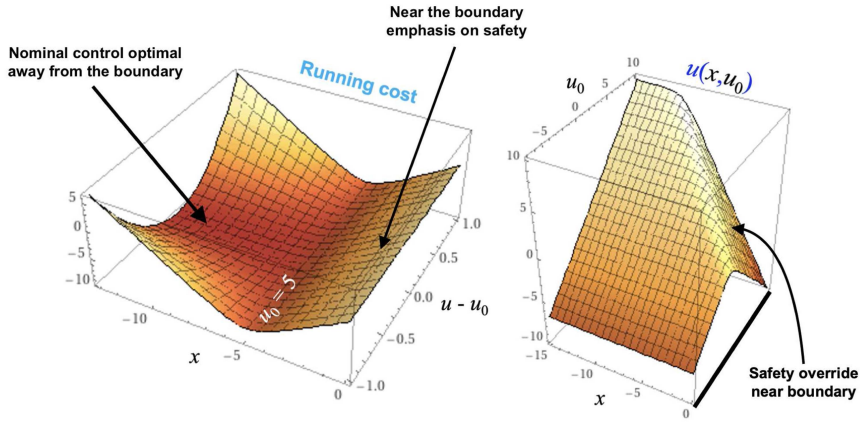

Fig. 1. Example in Section I-E. Left: running cost $\frac{-u_0+x+\sqrt{(u_0+x)^2+1}}{2}+\frac{1}{2}\frac{(u-u_0)^2}{u_0+x+\sqrt{(u_0+x)^2+1}}$ in (5) for $u_0=5$. Right: optimal feedback $u_{\rm S}^*=-x-\sqrt{(u_0+x)^2+1}$ in (4). Safe set $x\le 0$; safety boundary at $x=0$.

large. The inverse optimality results in Sections VIII and IX are given in the traditional mean sense, as in conventional stochastic optimal control.

In Section X, we consider parameter estimation and contrast the classical QP-based projection operator with a novel safety filter that, unlike projection, is continuous and also inverse optimal. In Section XI, we return to nonovershooting control [28], i.e., regulation to the safety boundary, beyond the strict-feedback systems in [28] and in the presence of deterministic or stochastic disturbances [30].

Notations: Let $a<0<b$. A continuous function $\gamma:(a,b)\to\mathbb{R}$ with $\gamma(0)=0$ is of extended class $\mathcal{K}_{(a,b)}$ if it is strictly increasing. A continuous function $\beta:(a,b)\times\mathbb{R}_{\ge 0}\to\mathbb{R}$ is of class $\mathcal{KL}_{(a,b)}$ if it is of class $\mathcal{K}_{(a,b)}$ in its first argument and has a zero limit as its second argument goes to infinity.

II. Input-to-State Safety (ISSf)

We start with definitions of a barrier function and safe set.

Definition 1: The scalar-valued differentiable function $h:\mathbb{R}^n\to\mathbb{R}$ with $\inf_{x\in\mathbb{R}^n}h(x)<0$ and $\sup_{x\in\mathbb{R}^n}h(x)>0$ is referred to as a *barrier function candidate*. The set $\mathcal{C}=\{x\in\mathbb{R}^n\mid h(x)\ge 0\}$ is referred to as a *safe set*.

Assumption 1: $\mathcal{C}$ is without isolated points.

Consider now the disturbance-driven system

$$\dot x=f(x)+g_1(x)d\,,\quad d\in\mathbb{R}^{m_1}\,. \tag{6}$$

Definition 2: The set $\mathcal{C}$ of the system (6) is said to be *input-to-state safe* (ISSf) if

$$h(x(t))\ge\beta(h(x_0),t)-\rho\left(\sup_{0\le\tau\le t}|d(\tau)|\right)\quad\forall t\ge 0\,. \tag{7}$$

where the function $\rho\in\mathcal{K}$ is referred to as the *ISSf gain function* and $\beta\in\mathcal{KL}_{(\inf h(\xi),\sup h(\xi))}=:\mathcal{KL}_h$.

This property is not new. Controller design ensuring ISSf, using backstepping for nonovershooting control, goes as far back as 2006 in the paper [28]—see the safety bound (61) of Theorem 3 with a disturbance of unlimited unknown bound $\bar d$, as well as the safety bound (90) of Proposition 1 with an observer-based nonovershooting controller.

The following definition is a very slightly adjusted version of [32, Def. 4].

Definition 3: The function h is called an *ISSf barrier function* (ISSf-BF) if there exist a function $\rho:[0,+\infty)\to[0,-\inf h(\xi))$ of class $\mathcal{K}$ and a function α in $\mathcal{K}_{(\inf h(\xi),\sup h(\xi))}$ such that, for all $x\in\mathbb{R}^n, d\in\mathbb{R}^{m_1}$,

$$\begin{aligned}&\min\{0,h(x)\}\le-\rho(|d|)\\ &\quad\Rightarrow\quad L_fh+L_{g_1}hd\ge-\alpha(h)\,.\end{aligned} \tag{8}$$

The following result is a variation on [32, Th. 1], proved by adapting [26, Th. 2.2] and [25, Th. 1].

Lemma 1: For the system (6), if there exists an ISSf-BF h such that (8) holds, then the system is ISSf with $\beta(r,t)$ in (7) defined by the solution to $\dot{\underline{h}}=-\alpha(\underline{h})$, $\underline{h}(0)=r$.

For converse barrier certificates, see [33].

III. ISSf-control barrier function (ISSf-CBFs)

Consider now, with loc. Lipschitz $f, g_1,$ and g_2, the system

$$\dot x=f(x)+g_1(x)d+g_2(x)u\,,\quad u\in\mathbb{R}^{m_2}\,. \tag{9}$$

Definition 4: A scalar differentiable function h is called an ISSf-CBF for (9) if there exists a class $\mathcal{K}$ function $\rho:\mathbb{R}_{\ge 0}\to[0,-\inf h(\xi))$ and $\alpha\in\mathcal{K}_{(\inf h(\xi),\sup h(\xi))}=:\mathcal{K}_h$ such that, for all $x\in\mathbb{R}^n, d\in\mathbb{R}^{m_1}$

$$\begin{aligned}&\min\{0,h(x)\}\le-\rho(|d|)\\ &\Rightarrow\quad\sup_{u\in\mathbb{R}^{m_2}}\{L_fh+L_{g_1}hd+L_{g_2}hu\}\ge-\alpha(h).\end{aligned} \tag{10}$$

The following result for CBFs is obtained by adapting our CLF result [27, Lemma 2.1].

Lemma 2: A pair (h,ρ) satisfies (10) if and only if

$$L_{g_2}h(x)=0\quad\Rightarrow\quad\omega(x)\ge 0 \tag{11}$$

where

$$\omega(x)=L_fh-|L_{g_1}h|\,\rho^{-1}(\max\{0,-h(x)\})+\alpha(h(x))\,. \tag{12}$$

ISSf-CBFs, which do not require the disturbance to be in a known compact set, are different from robust CBFs [12], [23].

That is the very purpose of the antecedent in the implication (10) and the term $\rho^{-1}(\max\{0,-h(x)\})$ in (12).

Theorem 1: If there exists an ISSf-CBF, the system (9) is rendered ISSf using the following Sontag-type control law[1]:

$$u = u_{\mathrm{S}}(x) = (L_{g_2}h)^{\mathrm{T}} \begin{cases} \kappa(x), & (L_{g_2}h)^T \neq 0 \\ 0, & (L_{g_2}h)^{\mathrm{T}} = 0 \end{cases} \tag{13}$$

where, with $\omega(x)$ defined in (12),

$$\begin{aligned} \kappa(x) &= \frac{-\omega + \sqrt{\omega^2 + (L_{g_2}h(L_{g_2}h)^{\mathrm{T}})^2}}{L_{g_2}h(L_{g_2}h)^{\mathrm{T}}} \\ &= \frac{L_{g_2}h(L_{g_2}h)^{\mathrm{T}}}{\omega + \sqrt{\omega^2 + (L_{g_2}h(L_{g_2}h)^{\mathrm{T}})^2}}\,. \end{aligned} \tag{14}$$

Proof: We substitute (13) into (9) and get

$$\begin{aligned} \dot{h} &= L_f h + L_{g_1}hd - \omega + \sqrt{\omega^2 + (L_{g_2}h(L_{g_2}h)^{\mathrm{T}})^2} \\ &\geq -\alpha(h(x)) + |L_{g_1}h|\left[\rho^{-1}(\max\{0,-h(x)\}) - |d|\right]. \end{aligned} \tag{15}$$

For $\min\{0,h(x)\} \leq -\rho(|d|)$, we, thus, have

$$\dot{h} = L_{f+g_2\alpha_{\mathrm{S}}} + L_{g_1}hd \geq -\alpha(h(x)) \tag{16}$$

which, thanks to Lemma 1, completes the proof of ISSf. □

IV. ISSf Filter

Now we turn our attention to the simultaneous objectives of maintaining safety and deviating as little as possible from the nominal $u_0(x,t)$. For that purpose, we rewrite (9) as

$$\dot{x} = f(x) + g_2(x)u_0 + g_1(x)d + g_2(x)(u-u_0)\,. \tag{17}$$

Let an ISSf-CBF $h(x)$ be available, with associated (ρ,α). Accounting for the inclusion of u_0 into the drift vector field (17), we modify (12) as

$$\begin{aligned} \omega(x,u_0) = L_{f+g_2u_0}h - |L_{g_1}h|\,\rho^{-1}(\max\{0,-h(x)\}) \\ + \alpha(h(x))\,. \end{aligned} \tag{18}$$

Then, we introduce the QP problem

$$\bar{u}_{\mathrm{QP}} = \arg\min_{v\in\mathbb{R}^{m_2}} |v|^2 \quad \text{subject to} \tag{19}$$

$$\omega(x,u_0) + L_{g_2}hv \geq 0\,. \tag{20}$$

The well-known explicit solution to this problem is [17]

$$\bar{u}_{\mathrm{QP}} = \begin{cases} 0, & \omega(x,u_0) \geq 0 \\ -\dfrac{\omega(x,u_0)}{|L_{g_2}h|^2}(L_{g_2}h)^T, & \omega(x,u_0) < 0\,. \end{cases} \tag{21}$$

Remark 1: Regarding the possible division by $L_{g_2}h = 0$ in the second case of (21), we recall that, by Lemma 2, every ISSf-CBF satisfies the implication $L_{g_2}h = 0 \Rightarrow \omega(x,u_0) \geq 0$, which is equivalent to the implication $\omega(x,u_0) < 0 \Rightarrow L_{g_2}h \neq 0$, and this precludes $L_{g_2}h$ being zero in the second case of (21), i.e., a division by zero is not possible.

While bounded, (21) is not necessarily continuous at points where $L_{g_2}h(x) = 0$. When the nominal u_0 is only a function of x, continuity can be ensured by assuming the following.

Assumption 2: For a given locally Lipschitz $u_0 : \mathbb{R}^n \to \mathbb{R}^{m_2}$, system (17) satisfies the *small control property* (SCP) [23], [27], [45], [47], i.e., there exists a (not necessarily known) continuous $\bar{u}_c(x)$ such that $L_{g_2}h(x) = 0 \Rightarrow \bar{u}_c(x) = 0$ and

$$\begin{aligned} \omega_c(x) = L_{f+g_2(u_0+\bar{u}_c)}h - |L_{g_1}h|\,\rho^{-1}(\max\{0,-h(x)\}) \\ + \alpha(h(x)) \geq 0\,. \end{aligned} \tag{22}$$

From (22) it follows, for ω defined in (18), that $\omega \leq 0 \Rightarrow |\omega| \leq |L_{g_2}h||\bar{u}_c|$, from which the continuity follows for $\bar{u}_{\mathrm{QP}}$ in (21), as well as for the Sontag controller in Theorem 1 (and 9). With the SCP, (21) and (13) are also locally Lipschitz on the open set $L_{g_2}h(x) \neq 0$.

With the QP safety filter (21), we have the following result.

Theorem 2: The control law

$$u = u_0 + \bar{u}_{\mathrm{QP}}(x,u_0) \tag{23}$$

with $\bar{u}_{\mathrm{QP}}(x,u_0)$ defined in (21) and $\omega(x,u_0)$ defined in (18) renders the system (17) ISSf with respect to the ISSf-CBF $h(x)$, with a gain function ρ, i.e., there exists $\beta \in \mathcal{KL}_h$ such that, for all $t \geq 0$

$$h(x(t)) \geq \beta(h(x_0),t) - \rho\left(\sup_{0\leq\tau\leq t}|d(\tau)|\right). \tag{24}$$

Proof: We substitute (23) and (21) into (9), get

$$\begin{aligned} \dot{h} &= L_{f+g_2u_0}h + L_{g_1}hd + L_{g_2}h\bar{u}_{\mathrm{QP}} \\ &= -\alpha(h(x)) + \omega + \max\{0,-\omega\} \\ &\quad + |L_{g_1}h|\rho^{-1}(\max\{0,-h(x)\}) + L_{g_1}hd \\ &\geq -\alpha(h(x)) + \max\{\omega,0\} \\ &\quad + |L_{g_1}h|\left[\rho^{-1}(\max\{0,-h(x)\}) - |d|\right] \\ &\geq -\alpha(h(x)) + |L_{g_1}h|\left[\rho^{-1}(\max\{0,-h(x)\}) - |d|\right] \end{aligned} \tag{25}$$

and invoke Lemma 1. □

Example 1: Consider the system

$$\dot{x} = u + (1+x^2)d \tag{26}$$

with an ISSf-CBF $h(x) = -x$. For some $\rho \in \mathcal{K}_\infty$, (18) is $\omega = -u_0 - (1+x^2)\rho^{-1}(\max\{0,x\}) + \alpha(h(x))$ and the QP formula (21) gives $\bar{u}_{\mathrm{QP}} = \min\{0, -u_0 - (1+x^2)\rho^{-1}(\max\{0,x\}) + \alpha(h(x))\}$. Taking, e.g., $\alpha(h) = h$, the overall feedback (23), given by $u = \min\{u_0, -(1+x^2)\rho^{-1}(\max\{0,x\}) - x\}$, guarantees, $\forall \rho \in \mathcal{K}_\infty$,

$$x(t) \leq \mathrm{e}^{-t}x_0 + \rho\left(\sup_{0\leq\tau\leq t}|d(\tau)|\right) \quad \forall t \geq 0\,. \tag{27}$$

□

Example 2: The safety filter (23) ensures safety but, on its own, does not guarantee forward completeness. For instance, the example in Section I-E, where $\dot{x} = u, h(x) = -x$, results in the QP safety filter $u = \min\{u_0, -x\}$. If the nominal control happens to be $u_0 = x^3$, the resulting overall control is $u = \min\{x^3, -x\}$. Within the safe set $x < 0$, this feedback becomes simply $u = x^3$ and gives a closed-loop system $\dot{x} = x^3$, which has a finite escape time. While this might at first disappoint, it should not. The safety filter ensures both safety and the exact conformity with the nominal control $u_0 = x^3$. If the user wishes to drive the solution to $-\infty$ in finite time, this is what the user gets with this safety filter. □

Hence, insisting on forward completeness may contradict the nominal objective and is not implied by safety. Nevertheless, for reasons of being able to state results like (24) for all $t \geq 0$, we

[1] See also the proof in [27, Th. 3.2] and [25, Remark 5].

seek conditions that ensure forward completeness. One way is to assume *unboundedness observability* [6].

Assumption 3: For system (9) with BF h and nominal u_0, there exists a proper, smooth $U : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ such that

$$L_{f+g_1 d+g_2 u} U \leq U + \sigma_1 \left(\max\{0, -h(x)\} + |d|\right) + \sigma_2(|u - u_0|) \tag{28}$$

for some $\sigma_1, \sigma_2 \in \mathcal{K}_\infty$ and for all $x \in \mathbb{R}^n, d \in \mathbb{R}^{m_1}$, and $u \in \mathbb{R}^{m_2}$. In addition, for some $M > 0$, the feedback u_c in Assumption 2 satisfies $|\bar{u}_c(x)| \leq \sigma_2^{-1}(MU(x))\ \forall x \in \mathbb{R}^n$.

With this assumption, we ensure that once we prove safety, namely, that $h(x(t)) \geq -\rho(\sup_{t\geq 0} |d(t)|)$ holds, it follows that $\max\{0, -h(x(t)) \leq \rho(\sup_{t\geq 0} |d(t)|)$ and, from (28) and (21), that $\dot{U} \leq (1+M)U + \sigma_1(\rho(\sup_{t\geq 0} |d(t)|) + \sup_{t\geq 0} |d(t)|)$, which implies forward completeness. The condition $|\bar{u}_c(x)| \leq \sigma_2^{-1}(U)$ in Assumption 3 is undoubtedly strong, but the alternative routes to ensuring the existence of solutions are even less appealing.

Due to limited space, in the rest of this article, we do not belabor regularity and existence issues. Assumptionssuch as Assumptions 2 and 3, can ensure these properties for all our safety filter designs, but at the expense of restricting u_0.

Remark 2: Since controls such as (21) appear in our article at least half a dozen times, for the sake of compactness, we write it as

$$\bar{u}_{\rm QP} = (L_{g_2} h)^T \frac{\max\{0, -\omega(x, u_0)\}}{|L_{g_2} h|^2} \tag{29}$$

with a recollection from (11) that $\omega < 0 \Rightarrow L_{g_2} h \neq 0$ and with a notational convention that $0/0 = 0$.

For the reader's future convenience, we point out that an alternative representation of (23) with the safety filter (21) is

$$u = \begin{cases} u_0, & \omega(x, u_0) \geq 0 \\ \chi_0(x) u_0 + \chi_1(x), & \omega(x, u_0) < 0 \end{cases} \tag{30}$$

where

$$\chi_0(x) = I - \frac{(L_{g_2} h)^T L_{g_2} h}{|L_{g_2} h|^2} \tag{31}$$

$$\chi_1(x) = -(L_{g_2} h)^T \frac{\omega_1(x)}{|L_{g_2} h|^2} \tag{32}$$

$$\omega_1(x) = L_f h - |L_{g_1} h|\, \rho^{-1}(\max\{0, -h(x)\}) + \alpha(h(x))\,. \tag{33}$$

The "half-Sontag" formula also generates min-norm control.

Theorem 3: The feedback

$$u = u_0 + \frac{1}{2} u_{\rm S} \tag{34}$$

with $u_{\rm S}$ defined in (13) and (14) and ω defined in (18) renders the system (17) ISSf and is the pointwise minimizer of $|v|^2$ subject to the following constraint more conservative than (20):

$$\frac{1}{2}\left(\omega - \sqrt{\omega^2 + \left(L_{g_2} h (L_{g_2} h)^{\rm T}\right)^2}\right) + L_{g_2} h v \geq 0\,. \tag{35}$$

Proof: The pointwise minimization result is immediate from (19) to (21). For (17), (34) ISSf follows from

$$\dot{h} = -\alpha(h(x)) + \frac{1}{2}\left(\omega + \sqrt{\omega^2 + \left(L_{g_2} h (L_{g_2} h)^{\rm T}\right)^2}\right) + |L_{g_1} h| \rho^{-1}(\max\{0, -h(x)\}) + L_{g_1} h d\,. \tag{36}$$

□

V. Inverse Optimal Assignment of ISSf Gain

In the system (17), there are two inputs: $u - u_0$ and d. This leads us to formulate the problem of safety filter design as a differential game [7], [8], for example, of the zero-sum type. In this game, the objective for both our control $u - u_0$ and for the disturbance d is for them to remain small. However, their objectives differ regarding safety: $u - u_0$ is tasked with keeping $h(x(t))$ from becoming too small, whereas d is tasked with making $h(x(t))$ small and, in fact, negative. Since our goal in designing $u - u_0$ is to make the ISSf gain function from d to the "safety violation" $-h(x)$ small, we refer to this problem as gain assignment.

We pursue the following zero-sum two-player minimax (supinf, to be precise) optimization problem:

$$\sup_{u-u_0 \in \mathcal{U}} \inf_{d \in \mathcal{D}} \left\{ \lim_{t\to\infty} \left[2\beta h(x(t)) + \int_0^t \Big(l(x, u_0) - (u - u_0)^{\rm T} R_2(x, u_0)(u - u_0) + \beta\lambda\gamma\left(\frac{|d|}{\lambda}\right)\Big) {\rm d}\tau \right] \right\} \tag{37}$$

where $\mathcal{U}$ and $\mathcal{D}$ are the sets of locally bounded functions of x. In this problem, $R_2(x, u_0) = R_2(x, u_0)^{\rm T} > 0$ for all x and u_0, γ and γ' are in class $\mathcal{K}_\infty$, the constants β and λ are positive, and $l(x, u_0)$ is a weight on the state, upper bounded by a class $\mathcal{K}_\infty$ function of h.

We do not approach the game (37) as a problem of direct determination of a Nash equilibrium but as an *inverse* problem: both the Nash control laws $u^*(x) - u_0$ and d^*, as well as the weights $l(x, u_0), R_2(x, u_0), \gamma(\cdot)$, are up to the designer to choose. Even $h(x)$ is available for design, for a given safe set $\mathcal{C}$.

Before we continue, let us introduce the following notation: For a class $\mathcal{K}_\infty$ function γ whose derivative exists and is also a class $\mathcal{K}_\infty$ function, $\ell\gamma$ denotes the Legendre–Fenchel transform

$$\ell\gamma(r) = \int_0^r (\gamma')^{-1}(s) ds \tag{38}$$

$$= r(\gamma')^{-1}(r) - \gamma\left((\gamma')^{-1}(r)\right)\,, \quad \text{(by Lemma 4.a)} \tag{39}$$

where $(\gamma')^{-1}(r)$ stands for the inverse function of $\dfrac{d\gamma(r)}{dr}$.

Theorem 4: Consider the auxiliary system of (9)

$$\dot{x} = f(x) - g_1(x)\ell\gamma(2|L_{g_1} h|)\frac{(L_{g_1} h(x))^{\rm T}}{|L_{g_1} h|^2} + g_2(x) u \tag{40}$$

with a nominal control law $u_0(x, t)$, where γ is a class $\mathcal{K}_\infty$ function whose derivative γ' is also a class $\mathcal{K}_\infty$ function. Suppose that, for a given u_0, there exists a matrix-valued function $R_2(x, u_0) = R_2(x, u_0)^{\rm T} > 0$ such that the control law of the form

$$u = u_0 + \bar{u}(x, u_0) := u_0 + R_2(x, u_0)^{-1} (L_{g_2} h)^{\rm T} \tag{41}$$

ensures safety of the system (40) with respect to CBF candidate $h(x)$, namely, ensures that

$$L_{f+g_2 u_0} h - \ell\gamma(2|L_{g_1} h|) + L_{g_2} h R_2^{-1} (L_{g_2} h)^{\rm T} \geq -\alpha(h(x)) \tag{42}$$

for some $\alpha \in \mathcal{K}_h$. Then, the control law

$$u = u_0 + \bar{u}^*(x, u_0) := u_0 + \beta\bar{u}(x, u_0) = u_0 + \beta R_2^{-1} (L_{g_2} h)^{\rm T}\,, \quad \beta \geq 2 \tag{43}$$

applied to (9) *maximizes* the cost functional

$$J(u-u_0) = \inf_{d\in\mathcal{D}}\Big\{\lim_{t\to\infty}\Big[2\beta h(x(t)) + \int_0^t \Big(l(x,u_0) - (u-u_0)^{\mathrm{T}} R_2(x,u_0)(u-u_0) + \beta\lambda\gamma\left(\frac{|d|}{\lambda}\right)\Big)\mathrm{d}\tau\Big]\Big\} \tag{44}$$

for any $\lambda \in (0,2]$, where

$$\begin{aligned} l(x,u_0) = & -2\beta\big[L_{f+g_2u_0}h - \ell\gamma(2|L_{g_1}h|) \\ & + L_{g_2}hR_2^{-1}\left(L_{g_2}h\right)^{\mathrm{T}}\big] \\ & - \beta(2-\lambda)\ell\gamma(2|L_{g_1}h|) \\ & - \beta(\beta-2)L_{g_2}hR_2^{-1}\left(L_{g_2}h\right)^{\mathrm{T}} \qquad (45) \\ \leq & \ 2\beta\alpha(h(x)) \qquad (46) \end{aligned}$$

is decrescent in the CBF h on the interval $(\inf h, \sup h)$.

Proof: Thanks to (42) and (45), we get (46). Substituting $l(x)$ into (44), it follows that

$$\begin{aligned} J(u) = & \inf_{d\in\mathcal{D}}\Big\{\lim_{t\to\infty}\Big[2\beta h(x(t)) \\ & + \int_0^t\Big(-2\beta L_{f+g_2u_0}h + \beta\lambda\ell\gamma(2|L_{g_1}h|) \\ & - \beta^2 L_{g_2}hR_2^{-1}\Big(L_{g_2}h\Big)^{\mathrm{T}} \\ & - (u-u_0)^{\mathrm{T}}R_2(u-u_0) + \beta\lambda\gamma\Big(\frac{|d|}{\lambda}\Big)\Big)d\tau\Big]\Big\} \\ = & \inf_{d\in\mathcal{D}}\Big\{\lim_{t\to\infty}\Big[2\beta h(x(t)) \\ & - 2\beta\int_0^t\Big(L_{f+g_2u_0}h + L_{g_1}hd + L_{g_2}h(u-u_0)\Big)d\tau \\ & - \int_0^t\Big((u-u_0)^{\mathrm{T}}R_2(u-u_0) - 2\beta L_{g_2}h(u-u_0) \\ & + \beta^2 L_{g_2}hR_2^{-1}\Big(L_{g_2}h\Big)^{\mathrm{T}}\Big)d\tau \\ & + \int_0^t\Big(\beta\lambda\gamma\Big(\frac{|d|}{\lambda}\Big) + 2\beta L_{g_1}hd \\ & + \beta\lambda\ell\gamma(2|L_{g_1}h|)\Big)d\tau\Big]\Big\} \\ = & \inf_{d\in\mathcal{D}}\Big\{\lim_{t\to\infty}\Big[2\beta h(x(t)) - 2\beta\int_0^t dh \\ & - \int_0^t (u-u_0-\bar{u}^*)^{\mathrm{T}}R_2(u-u_0-\bar{u}^*)d\tau \\ & + \beta\int_0^t\Big[\lambda\gamma\Big(\frac{|d|}{\lambda}\Big) - \lambda\gamma\Big((\gamma')^{-1}(2|L_{g_1}h|)\Big) \\ & + 2\Big(\lambda|L_{g_1}h|(\gamma')^{-1}(2|L_{g_1}h|) + L_{g_1}hd\Big)\Big]d\tau\Big]\Big\} \end{aligned}$$

(by (39))

$$\begin{aligned} & = 2\beta h(x(0)) + \beta\lambda\inf_{d\in\mathcal{D}}\int_0^\infty \Pi(d,d^*)\,\mathrm{d}t \\ & \quad - \int_0^\infty (u-u_0-\bar{u}^*)^{\mathrm{T}}R_2(u-u_0-\bar{u}^*)d\tau\,\mathrm{d}t \qquad (47) \end{aligned}$$

where

$$\begin{aligned} \Pi(d,d^*) = & \ \gamma\Big(\frac{|d|}{\lambda}\Big) - \gamma\Big(\frac{|d^*|}{\lambda}\Big) \\ & - \gamma'\Big(\frac{|d^*|}{\lambda}\Big)\frac{(d^*)^{\mathrm{T}}}{\lambda|d^*|}(d^*-d) \qquad (48) \end{aligned}$$

and

$$d^*(x) = -\lambda(\gamma')^{-1}(2|L_{g_1}h|)\frac{(L_{g_1}h)^{\mathrm{T}}}{|L_{g_1}h|}\,. \tag{49}$$

By Lemma 4.d, $\Pi(d,d^*)$ can be rewritten as

$$\begin{aligned} \Pi(d,d^*) = & \ \gamma\left(\frac{|d|}{\lambda}\right) + \ell\gamma\left(\gamma'\left(\frac{|d^*|}{\lambda}\right)\right) \\ & + \gamma'\left(\frac{|d^*|}{\lambda}\right)\frac{(d^*)^{\mathrm{T}}}{|d^*|}\frac{d}{\lambda}\,. \qquad (50) \end{aligned}$$

Then, by Lemma 5, we have

$$\begin{aligned} \Pi(d,d^*) \geq & \ \gamma\left(\frac{|d|}{\lambda}\right) + \ell\gamma\left(\gamma'\left(\frac{|d^*|}{\lambda}\right)\right) \\ & - \gamma\left(\frac{|d|}{\lambda}\right) - \ell\gamma\left(\gamma'\left(\frac{|d^*|}{\lambda}\right)\right) \\ = & \ 0 \qquad (51) \end{aligned}$$

and $\Pi(d,d^*) = 0$ if and only if $\frac{d}{\lambda} = (\gamma')^{-1}\left(\gamma'\left(\frac{|d^*|}{\lambda}\right)\right)\frac{d^*}{|d^*|}$, that is

$$\Pi(d,d^*) = 0 \quad \text{iff} \quad d = d^*\,. \tag{52}$$

Thus,

$$\inf_{d\in\mathcal{D}}\int_0^\infty \Pi(d,d^*)\,\mathrm{d}t = 0 \tag{53}$$

and the "worst-case" disturbance is given by (49). The maximum of (47) is reached with $u = u_0 + \bar{u}^*$. Hence, the control law (43) maximizes the cost functional (44). The value function of (44) is $J^*(x) = 2\beta h(x)$. □

The parameter $\beta \geq 2$ in the statement of Theorem 4 represents a design degree of freedom. The parameter λ (note that it parameterizes not only the penalty on the disturbance but also the penalty on the state's proximity to the boundary, i.e., the reward for the state's distance from the boundary, $l(x,u_0)$) indicates that the same control law is inverse optimal with respect to an entire family of different cost functionals.

Remark 3: One approach to studying safety in the presence of inputs and disturbances is reachability [9], [12], [56], [57], where Hamilton–Jacobi–Isaacs (HJI) PDEs arise and need to be solved. Even though not explicit in the proof of Theorem 4, the CBF $h(x)$ solves the following family of HJI equations:

$$\begin{aligned} & L_{f+g_1u_0}h + \frac{\beta}{2}L_{g_2}hR_2(x,u_0)^{-1}\left(L_{g_2}h\right)^{\mathrm{T}} - \frac{\lambda}{2}\ell\gamma(2|L_{g_1}h|) \\ & \quad + \frac{l(x,u_0)}{2\beta} = 0 \qquad (54) \end{aligned}$$

parameterized by $(\beta,\lambda) \in [2,\infty)\times(0,2]$. □

Remark 4: It is also easily seen from the proof of Theorem 4 that, even for initial conditions on the boundary, the achieved

disturbance attenuation level is

$$2\beta h(x(t))+2\beta\int_0^\infty \alpha(h(x))\,\mathrm{d}t \geq 2\beta h(x(t))+\int_0^\infty l(x,u_0)\,\mathrm{d}t$$
$$\geq \int_0^\infty (u-u_0)^{\mathrm{T}} R_2(x,u_0)(u-u_0)\,\mathrm{d}t-\beta\lambda\int_0^\infty \gamma\left(\frac{|d|}{\lambda}\right)\,\mathrm{d}t$$
$$\geq -\beta\lambda\int_0^\infty \gamma\left(\frac{|d|}{\lambda}\right)\,\mathrm{d}t\,. \tag{55}$$

Summarizing, we refer to the property

$$h(x(t))+\int_0^\infty \alpha(h(x))\,\mathrm{d}t \geq -\frac{\lambda}{2}\int_0^\infty \gamma\left(\frac{|d|}{\lambda}\right)\,\mathrm{d}t \tag{56}$$

as integral *ISSf.* □

Example 3: Consider the system from Example 1. Take $\gamma(r)=\ell\gamma(2r)=r^2$. With

$$R_2=\frac{1}{\max\{0,u_0-\alpha(h(x))\}+(1+x^2)^2}>0 \tag{57}$$

condition (42) is satisfied. The control (43) is given by

$$u=u_0-\frac{\beta}{R_2}=u_0+\beta\left[\min\{0,-u_0+\alpha(h(x))\}-(1+x^2)^2\right] \tag{58}$$

and, for all $\beta\geq 2$, is the maximizer of

$$J(u-u_0)=\inf_{d\in\mathcal{D}}\left\{\lim_{t\to\infty}\left[2\beta x(t)+\int_0^t\left(l(x,u_0)\right.\right.\right.$$
$$\left.\left.\left.-R_2(u-u_0)^2+\frac{\beta}{\lambda}d^2\right)\mathrm{d}\tau\right]\right\} \tag{59}$$

for any $\lambda\in(0,2]$, with $l(x,u_0)\leq -2\beta x$, and achieves

$$x(+\infty)+\int_0^\infty x(t)\,\mathrm{d}t\leq\frac{1}{2\lambda}\int_0^\infty d^2(t)\,\mathrm{d}t \tag{60}$$

and, $\forall\beta\geq 1$, controller (58) with $\alpha(h)=h$ guarantees

$$x(t)\leq \mathrm{e}^{-t}x_0+\frac{1}{4}\left(\sup_{0\leq\tau\leq t}|d(\tau)|\right)^2\quad\forall t\geq 0\,. \tag{61}$$

□

Following the general result in Theorem 4, a natural question arises: Is the ISSf QP safety filter (21), (18) inverse optimal? The following theorem, proven similarly to Theorem 4, answers the question in the affirmative.

Theorem 5: Consider system (17) with associated ISSf-CBF h and a gain function ρ. For any $\beta\geq 2$, the control law

$$u=u_0+\bar{u}^*_{\mathrm{QP}}(x,u_0)=u_0+\beta\bar{u}_{\mathrm{QP}}(x,u_0) \tag{62}$$

with $\bar{u}_{\mathrm{QP}}$ defined in (21) and ω defined in (18), *maximizes*

$$J(u-u_0)=\inf_{d\in\mathcal{D}}\left\{\lim_{t\to\infty}\left[2\beta h(x(t))+\int_0^t\left(l(x,u_0)\right.\right.\right.$$
$$\left.\left.\left.-R_2(x,u_0)|u-u_0|^2+\frac{\beta}{\lambda}R_1(x)|d|^2\right)\mathrm{d}\tau\right]\right\} \tag{63}$$

for all $\lambda\in(0,2]$, where

$$R_1(x)=\frac{1}{\rho^{-1}(\max\{0,-h(x)\})}>0 \tag{64}$$

$$R_2(x,u_0)=\frac{|L_{g_2}h|^2}{\max\{0,-\omega\}}>0 \tag{65}$$

$$l(x,u_0)\leq 2\beta\alpha(h(x))\,. \tag{66}$$

The weight R_1 in (64) is infinite in the safe set $h(x)\geq 0$ where the "optimal disturbance"

$$d^*(x)=-\lambda\rho^{-1}(\max\{0,-h(x)\})\frac{(L_{g_1}h(x))^{\mathrm{T}}}{|L_{g_1}h|} \tag{67}$$

spends no effort. Likewise, R_2 in (65) is infinite when $\omega\geq 0$ since control (21) puts in no effort when u_0 makes the system safe on its own. We also recall from Remark 1 that (21) precludes $L_{g_2}h$ from being zero when $\omega<0$, so R_2 can, in fact, never be zero, namely, $u-u_0$ is penalized for all x.

Example 4: Back to Example 1, control $u=u_0+\beta\bar{u}_{\mathrm{QP}}$, $\beta\geq 2$, with $\bar{u}_{\mathrm{QP}}=\min\{0,-u_0-(1+x^2)\rho^{-1}(\max\{0,x\})+x\}$, results in

$$x(+\infty)+\int_0^\infty x(t)\,\mathrm{d}t\leq\frac{1}{2\lambda}\int_0^\infty\frac{d^2(t)}{\rho^{-1}(\max\{0,x(t)\})}\,\mathrm{d}t \tag{68}$$

which, unlike control (58) in Example 3, fails to achieve a finite integral gain in the safe set $x\leq 0$ like (60). □

VI. Inverse Optimal QP Safety Filter for Disturbance-Free Systems

In the disturbance-free system,

$$\dot{x}=f(x)+g_2(x)u_0+g_2(x)(u-u_0) \tag{69}$$

let us introduce

$$\omega_2(x,u_0)=L_{f+g_2u_0}h+\alpha(h(x)) \tag{70}$$

and the control law $u=u_0+\bar{u}_{\mathrm{QP2}}(x,u_0)$ with (recalling Remark 2 on division by $L_{g_2}h=0$)

$$\bar{u}_{\mathrm{QP2}}=(L_{g_2}h)^T\frac{\max\{0,-\omega_2(x,u_0)\}}{|L_{g_2}h|^2}\,. \tag{71}$$

This standard QP safety filter renders system (69) safe w.r.t. CBF $h(x)$. The next result follows from Theorem 5.

Corollary 1: For system (69) and any $\beta\geq 2$, the control

$$u=u_0+\bar{u}^*_{\mathrm{QP2}}(x,u_0)=u_0+\beta\bar{u}_{\mathrm{QP2}}(x,u_0) \tag{72}$$

with $\bar{u}_{\mathrm{QP2}}$ defined in (71), *maximizes* the cost functional

$$J(u-u_0)=\lim_{t\to\infty}\left[2\beta h(x(t))+\int_0^t\left(l(x,u_0)\right.\right.$$
$$\left.\left.-\frac{|L_{g_2}h|^2|u-u_0|^2}{\max\{0,-\omega_2\}}\right)\mathrm{d}\tau\right] \tag{73}$$

with $l(x,u_0)\leq 2\beta\alpha(h)$.

It is from this corollary that the illustrations in Section I-E follow, along with the intuition provided there. By examining the payoff (73), one notes that the deviation $u-u_0$ is being minimized, whereas the "terminal safety" payoff $\beta(h(x(+\infty)))$ and the "running safety" payoff $l(x,u_0)$ are being maximized. The running safety payoff is meaningful because $l(x,u_0)\leq 2\beta\alpha(h)$: if the payoff $l(x,u_0)$ is high, then its upper bounding explicit safety payoff $\alpha(h)$ is certainly high.

In some application areas (computer science, subfields of robotics, etc.), RCBFs of the form $B(x)=1/h(x)$ are preferred. We return to the QP filter $u=u_0+\bar{u}_{\mathrm{QP2}}$ with (71) and (70), which can be expressed as

$$u=u_0+\tilde{u}_{\mathrm{QP}}(x,u_0) \tag{74}$$

with (recalling Remark 2 on division by $L_gB=0$)

$$\tilde{u}_{\mathrm{QP}}=-(L_gB)^T\frac{\max\{0,\tilde{\omega}(x,u_0)\}}{|L_gB|^2} \tag{75}$$

$$\tilde{\omega}(x,u_0) = L_{f+gu_0}B - \bar{\alpha}\left(\frac{1}{B(x)}\right) \tag{76}$$

for some $\bar{\alpha} \in \mathcal{K}$, and recast Corollary 1 for RCBFs.

Corollary 2: Let a function B be such that $h = 1/B$ satisfies Definition 1, and let B be an RCBF, namely, let B satisfy the condition $L_gB = 0 \Rightarrow \tilde{\omega} \leq 0$, with $\tilde{\omega}$ defined in (76). Then, the safety filter

$$u = u_0 + \beta\tilde{u}_{\rm QP}(x,u_0)\,, \quad \beta \geq 2 \tag{77}$$

with $\tilde{u}_{\rm QP}$ defined in (75), *minimizes*

$$J(u-u_0) = \lim_{t\to\infty}\Bigg[-\frac{2\beta}{B(x(t))} + \int_0^t \Bigg(\bar{l}(x,u_0) + \frac{|L_gB|^2|u-u_0|^2}{\max\{0,\tilde{\omega}\}}\Bigg)\mathrm{d}\tau\Bigg] \tag{78}$$

with $\bar{l}(x,u_0) \geq -\frac{2\beta}{B^2(x(t))}\bar{\alpha}(\frac{1}{B(x)})$.

VII. Stochastic CBFs

We return to the barrier function in Definition 1, under Assumption 1, but now consider the stochastic system

$$\mathrm{d}x = f(x)\,\mathrm{d}t + g_1(x)\,\mathrm{d}w \tag{79}$$

where w is an r-dimensional standard Wiener process, and f and g are locally Lipschitz.

For the barrier function candidate $h(x)$, we recall that Itô's lemma states that

$$\mathrm{d}h = \mathcal{L}h\,\mathrm{d}t + L_{g_1}h\,\mathrm{d}w \tag{80}$$

where

$$\mathcal{L}h = L_fh + \frac{1}{2}\mathrm{Tr}\left\{g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\right\} \tag{81}$$

is referred to as the *infinitesimal generator* of h.

We say that the system (79) satisfies the *stochastic barrier function condition* (SBFc) if there exists a function $\alpha \in \mathcal{K}_h$ such that, for all $x \in \mathbb{R}^n$, the following function is nonnegative:

$$\omega(x) = L_fV + \frac{1}{2}\mathrm{Tr}\left\{g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\right\} + \alpha(h)\,. \tag{82}$$

From here on, we proceed formally, with systems and controllers that satisfy the SBFc, without going a step further to establish safety in probability, or at least in the mean, which would be done by employing the techniques as in the proof given in [26, Th. 3.2], the techniques in [13, Th. 3], or the technique in the proof given in [30, Lemma 1].

Now we turn our attention to systems that, in addition to the noise input w, have a control input $u \in \mathbb{R}^{m_2}$

$$\mathrm{d}x = f(x)\,\mathrm{d}t + g_1(x)\,\mathrm{d}w + g_2(x)u\,\mathrm{d}t\,. \tag{83}$$

Definition 5: A scalar differentiable function h is called a *stochastic control barrier function (SCBF)* for (83) if there exists a function $\alpha \in \mathcal{K}_{(0,\sup h(\xi))}$ such that the following implication holds for all $x \in \{\eta \in \mathbb{R}^n | 0 \leq h(\eta) < \sup_{\xi\in\mathbb{R}^n} h(\xi)\}$:

$$\sup_{u\in\mathbb{R}^{m_2}}\left\{L_fV + \frac{1}{2}\mathrm{Tr}\left\{g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\right\} + L_{g_2}hu\right\} \geq -\alpha(h). \tag{84}$$

The following is obtained by adapting [27, Lemma 2.1].

Lemma 3: A function h is an SCBF, namely, it satisfies (84) in Definition 5, if and only if

$$L_{g_2}h = 0 \quad \Rightarrow \quad \omega \geq 0 \tag{85}$$

where $\omega(x)$ is defined in (82).

Next, a Sontag-type control law ensures stochastic safety.

Theorem 6: Under the control law (13), (14) with $\omega(x)$ defined in (82), the system (83) satisfies the SBFc.

Proof: A direct substitution yields

$$\mathcal{L}h = L_fh + \frac{1}{2}\mathrm{Tr}\left\{g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\right\} + L_{g_2}hu_{\rm S} \geq -\alpha(h(x))\,. \tag{86}$$

□

VIII. Inverse Optimal Stochastic Safety Filters

Next, we turn our attention to systems where, in addition to ensuring safety in the presence of noise w, the task of control input u is to stick close to the nominal u_0

$$\mathrm{d}x = [f(x) + g_2(x)u_0]\,\mathrm{d}t + g_1(x)\,\mathrm{d}w + g_2(x)(u-u_0)\,\mathrm{d}t\,. \tag{87}$$

Theorem 7: Consider the control law

$$u = u_0 + \bar{u}(x,u_0) \tag{88}$$

$$\bar{u}(x,u_0) = R_2^{-1}\left(L_{g_2}h\right)^{\rm T}\frac{\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)}{\left|L_{g_2}hR_2^{-1/2}\right|^2} \tag{89}$$

where γ_2 is a class $\mathcal{K}_\infty$ function whose derivative is also a class $\mathcal{K}_\infty$ function, and $R_2(x)$ is a matrix-valued function such that $R_2(x) = R_2(x)^{\rm T} > 0$. If the control law (89) makes the system (87) satisfy the SBFc with respect to an SCBF candidate $h(x)$, namely, if the following condition holds:

$$L_{f+g_2u_0}h + \frac{1}{2}\mathrm{Tr}\left\{g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\right\} + \ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right) \geq -\alpha(h) \tag{90}$$

then the control law

$$u = u_0 + \bar{u}^*(x,u_0) \tag{91}$$

$$\bar{u}^* = \frac{\beta}{2}R_2^{-1}\left(L_{g_2}h\right)^{\rm T}\frac{(\gamma_2')^{-1}\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)}{\left|L_{g_2}hR_2^{-1/2}\right|}, \quad \beta \geq 2 \tag{92}$$

also makes the system (87) satisfy the SBFc and, moreover, *maximizes* the cost functional

$$J(u-u_0) = \lim_{t\to\infty} E\Bigg\{2\beta h(x(t)) + \int_0^t \Bigg[l(x,u_0) - \beta^2\gamma_2\left(\frac{2}{\beta}\left|R_2^{1/2}(u-u_0)\right|\right)\Bigg]d\tau\Bigg\} \tag{93}$$

where

$$\begin{aligned} l(x,u_0) = &-2\beta\Bigg[L_{f+g_2u_0}h + \frac{1}{2}\mathrm{Tr}\left\{g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\right\} \\ &+ \ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)\Bigg] \\ &- \beta(\beta-2)\,\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right) \\ \leq\ & 2\beta\alpha(h)\,. \end{aligned} \tag{94}$$

Proof: Before we engage into proving that the control law (92) maximizes (93), we first show that makes the system (87)

satisfy the SBFc. With Lemma 4 we get

$$\begin{aligned}\mathcal{L}h\mid_{(92)} &= L_{f+g_2u_0}h+\frac{1}{2}\mathrm{Tr}\left\{g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right\}\\ &\quad+\frac{\beta}{2}\left|L_{g_2}hR_2^{-1/2}\right|(\gamma_2')^{-1}\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)\\ &= L_{f+g_2u_0}h+\frac{1}{2}\mathrm{Tr}\left\{g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right\}\\ &\quad+\frac{\beta}{2}\left[\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)\right.\\ &\quad\left.+\gamma_2\left((\gamma_2')^{-1}\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)\right)\right]\\ &\geq \mathcal{L}h\mid_{(89)}\geq -\alpha(h)\end{aligned}\tag{95}$$

which proves that (92) makes the system (87) satisfy the SBFc.

Now we prove optimality. Recalling that the Itô differential of h is

$$\mathrm{d}h=\mathcal{L}h(x)\,\mathrm{d}t+\frac{\partial h}{\partial x}g_1(x)\,\mathrm{d}w\tag{96}$$

according to the property of Itô's integral [36, Th. 3.9], we get

$$E\left\{h(0)-h(t)+\int_0^t\mathcal{L}h(x(\tau))d\tau\right\}=0.\tag{97}$$

Then, substituting $l(x)$ into $J(u)$, we have

$$\begin{aligned}J(u-u_0) &= \lim_{t\to\infty}E\left\{2\beta h(x(t))+\int_0^t\left[l(x,u_0)\right.\right.\\ &\quad\left.\left.-\beta^2\gamma_2\left(\frac{2}{\beta}\left|R_2^{1/2}(u-u_0)\right|\right)\right]d\tau\right\}\\ &= 2\beta E\left\{h(x(0))\right\}+\lim_{t\to\infty}E\left\{\int_0^t\left[2\beta\mathcal{L}h\mid_{(87)}\right.\right.\\ &\quad\left.\left.+l(x,u_0)-\beta^2\gamma_2\left(\frac{2}{\beta}\left|R_2^{1/2}(u-u_0)\right|\right)\right]d\tau\right\}\\ &= 2\beta E\left\{h(x(0))\right\}\\ &\quad+\lim_{t\to\infty}E\left\{\int_0^t\left[-\beta^2\gamma_2\left(\frac{2}{\beta}\left|R_2^{1/2}(u-u_0)\right|\right)\right.\right.\\ &\quad\left.\left.-\beta^2\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)+2\beta L_{g_2}h(u-u_0)\right]d\tau\right\}.\end{aligned}\tag{98}$$

Now we note that

$$\gamma_2'\left(\frac{2}{\beta}\left|R_2^{1/2}\bar{u}^*\right|\right)=\left|L_{g_2}hR_2^{-1/2}\right|\tag{99}$$

which yields

$$\begin{aligned}&J(u-u_0)\\ &=\lim_{t\to\infty}E\left\{\int_0^t\left[-\beta^2\gamma_2\left(\left|\frac{2}{\beta}R_2^{1/2}(u-u_0)\right|\right)\right.\right.\\ &\quad\left.\left.-\beta^2\ell\gamma_2\left(\gamma_2'\left(\left|\frac{2}{\beta}R_2^{1/2}\bar{u}^*\right|\right)\right)\bar{u}^*\right.\right.\\ &\quad\left.\left.+2\beta\gamma_2'\left(\left|\frac{2}{\beta}R_2^{1/2}\bar{u}^*\right|\right)\frac{\left(\frac{2}{\beta}R_2^{1/2}\bar{u}^*\right)^{\mathrm{T}}}{\left|\frac{2}{\beta}R_2^{1/2}\bar{u}^*\right|}R_2^{1/2}(u-u_0)\right]d\tau\right\}\\ &\quad+2\beta E\left\{h(x(0))\right\}.\end{aligned}\tag{100}$$

With the general Young inequality (Lemma 5), we obtain

$$\begin{aligned}J(u-u_0) &\leq 2\beta E\left\{h(x(0))\right\}+\lim_{t\to\infty}E\\ &\quad\times\left\{\int_0^t\left[-\beta^2\gamma_2\left(\left|\frac{2}{\beta}R_2^{1/2}(u-u_0)\right|\right)\right.\right.\\ &\quad-\beta^2\ell\gamma_2\left(\gamma_2'\left(\left|\frac{2}{\beta}R_2^{1/2}\bar{u}^*\right|\right)\right)\\ &\quad+\beta^2\gamma_2\left(\left|\frac{2}{\beta}R_2^{1/2}(u-u_0)\right|\right)\\ &\quad\left.\left.+\beta^2\ell\gamma_2\left(\gamma_2'\left(\left|\frac{2}{\beta}R_2^{1/2}\bar{u}^*\right|\right)\right)\right]d\tau\right\}\\ &= 2\beta E\left\{h(x(0))\right\}\end{aligned}\tag{101}$$

where the equality holds if and only if

$$\begin{aligned}&\gamma_2'\left(\left|\frac{2}{\beta}R_2^{1/2}\bar{u}^*\right|\right)\frac{\left(\frac{2}{\beta}R_2^{1/2}\bar{u}^*\right)^{\mathrm{T}}}{\left|\frac{2}{\beta}R_2^{1/2}\bar{u}^*\right|}\\ &=\gamma_2'\left(\left|\frac{2}{\beta}R_2^{1/2}(u-u_0)\right|\right)\frac{\left(\frac{2}{\beta}R_2^{1/2}(u-u_0)\right)^{\mathrm{T}}}{\left|\frac{2}{\beta}R_2^{1/2}(u-u_0)\right|}\end{aligned}\tag{102}$$

that is, when $u-u_0=\bar{u}^*$. Thus,

$$\arg\max_{u-u_0}J(u-u_0)=\bar{u}^*\tag{103}$$

$$\max_{u-u_0}J(u-u_0)=2\beta E\left\{h(x(0))\right\}.\tag{104}$$

□

Remark 5: Similar to Remark 3, even though not explicit in the proof of Theorem 7, $h(x)$ solves the following family of *Hamilton-Jacobi-Bellman* equations parameterized by $\beta\in[2,\infty)$:

$$\begin{aligned}&L_{f+g_2u_0}h+\frac{1}{2}\mathrm{Tr}\left\{g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right\}+\frac{\beta}{2}\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)\\ &\quad+\frac{l(x,u_0)}{2\beta}=0.\end{aligned}\tag{105}$$

□

Theorem 7 establishes inverse optimality but does not design a controller that meets condition (90). We pursue an inverse optimal safety-ensuring control next, using QP.

Theorem 8: For the system (87) and for any $\beta\geq 2$, the control law

$$u=u_0+\bar{u}^*_{\mathrm{QP2}}(x,u_0)=u_0+\beta\bar{u}_{\mathrm{QP2}}(x,u_0)\tag{106}$$

employing a standard QP safety filter

$$\bar{u}_{\mathrm{QP2}}=(L_{g_2}h)^T\frac{\max\left\{0,-\omega(x,u_0)\right\}}{|L_{g_2}h|^2}\tag{107}$$

with

$$\omega_2(x,u_0)=L_{f+g_2u_0}h+\frac{1}{2}\mathrm{Tr}\left\{g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right\}+\alpha(h(x))\tag{108}$$

maximizes the cost functional

$$J(u-u_0)=\lim_{t\to\infty}E\Bigg\{2\beta h(x(t))+\int_0^t\Bigg(l(x,u_0)-\frac{|L_{g_2}h|^2|u-u_0|^2}{\max\{0,-\omega_2\}}\Bigg)\mathrm{d}\tau\Bigg\} \tag{109}$$

for any $\lambda\in(0,2]$ and with some $l(x,u_0)\leq 2\beta\alpha(h)$.

Proof: By verifying that (90) is met with $\gamma_2(r)=\frac{1}{4}r^2$. □

Example 5: While a stochastic disturbance does not always have a detrimental effect on the SBFc condition, we construct an example in which the stochastic effect is indeed detrimental and where a QP safety filter acts to mitigate this effect. Consider the system

$$\mathrm{d}x=u\,\mathrm{d}t+(1-x)\,\mathrm{d}w\,. \tag{110}$$

Let us take $u_0=0$, $h(x)=\ln(1-x)$, and $\alpha(h)=h$. We obtain that (108) gives $\omega_2(x,u_0)=-\frac{1}{2}+\ln(1-x)$, which yields (107) in the following form:

$$\bar{u}_{\mathrm{QP2}}=(x-1)\max\left\{0,\frac{1}{2}-\ln(1-x)\right\}. \tag{111}$$

The safety filter kicks in when $x>1-\sqrt{\mathrm{e}}$, which is negative. If the stochastic disturbance w was absent, the control would be $\bar{u}_{\mathrm{QP2}}=(x-1)\max\{0,-\ln(1-x)\}$ and the safety filter would kick in only at $x=0$, i.e., never if $x_0<0$.

IX. Noise-to-State Safety Filters

In Sections VII and VIII, we studied stochastic systems of the form (83) with a known, unity covariance. This is quite limiting, regardless of the unity-intensity noise being the standard in stochastic optimal control. A stochastic disturbance acting on a system may be of unknown and time-varying incremental covariance $\Sigma(t)\Sigma(t)^{\mathrm{T}}\,\mathrm{d}t$, i.e.,

$$E\left\{\mathrm{d}w\,\mathrm{d}w^{\mathrm{T}}\right\}=\Sigma(t)\Sigma(t)^{\mathrm{T}}\,\mathrm{d}t \tag{112}$$

where $\Sigma(t)$ is a bounded function taking values in the set of non-negative definite matrices. For matrices $X=[x_1,x_2,\ldots,x_n]$, we use the Frobenius norm

$$|X|_{\mathcal{F}}\triangleq\left(\mathrm{Tr}\left\{X^{\mathrm{T}}X\right\}\right)^{1/2}=\left(\mathrm{Tr}\left\{XX^{\mathrm{T}}\right\}\right)^{1/2} \tag{113}$$

and note that $|X|_{\mathcal{F}}=|\mathrm{col}(X)|$, where $\mathrm{col}(X)=[x_1^{\mathrm{T}},x_2^{\mathrm{T}},\ldots,x_n^{\mathrm{T}}]^{\mathrm{T}}$.

When the covariance is unknown and time-varying, it needs to be treated as a deterministic disturbance in Sections II–IV. Accordingly, only a graceful degradation of safety in the presence of the disturbance $\Sigma(t)$ can be expected, as in (7). We refer to such a stochastic property as *noise-to-state safety* (NSSf). However, we do not conduct analysis of achieving such a property in probability or in the mean. We just pursue the attainment of the following condition:

$$\begin{aligned}&\min\{0,h(x)\}\leq-\rho\left(\left|\Sigma\Sigma^{\mathrm{T}}\right|_{\mathcal{F}}\right)\\&\Downarrow\\&L_fh+L_{g_1}u+\tfrac{1}{2}\mathrm{Tr}\left\{\Sigma^{\mathrm{T}}g^{\mathrm{T}}\tfrac{\partial^2h}{\partial x^2}g\Sigma\right\}\geq-\alpha(h)\end{aligned} \tag{114}$$

for system (83) with (112), by feedback $u=u_0+\bar{u}(x,u_0)$ for a nominal control law u_0, and call this condition the *noise-to state barrier function condition* (NSBFc).

Heretofore, we have dealt with CBF and ISSf-CBFs. We say that a function h is a *noise-to-state safety control barrier function* (NSSf-CBF) if, in addition to its usual conditions, it satisfies the implication

$$L_{g_2}h=0\quad\Rightarrow\quad\omega\geq0 \tag{115}$$

where

$$\omega(x,u_0)=L_{f+g_2u_0}h(x)+\alpha(h)-\frac{1}{2}\left|g_1^{\mathrm{T}}\frac{\partial^2h}{\partial x^2}g_1\right|_{\mathcal{F}}\rho^{-1}(\max\{0,-h(x)\}) \tag{116}$$

for a class $\mathcal{K}$ $\rho:[0,+\infty)\to[0,-\inf h(\xi))$ and $\alpha\in\mathcal{K}_h$.

Theorem 9: Under either the Sontag-type control (13), (14), or the QP control (21), along with $\omega(x,u_0)$ defined in (116), the system (83) with (112) satisfies the NSBFc in (114).

Proof: For both control laws, a direct substitution yields $\mathcal{L}h\geq-\alpha(h(x))$ whenever $\min\{0,h(x)\}\leq-\rho(|\Sigma\Sigma^{\mathrm{T}}|_{\mathcal{F}})$. □

Example 6: To illustrate a design for NSSf, we return to Example 1 but with the disturbance d replaced by white noise of unknown variance $\sigma(t)$, namely, to

$$\mathrm{d}x=u\,\mathrm{d}t+(1+x^2)\sigma(t)\,\mathrm{d}w\,. \tag{117}$$

To vary the design a bit but still keep it simple, we choose $h(x)=-x^3$ and $\alpha(h)=3\,h$. Conducting the calculations with (116), with arbitrary $\rho\in\mathcal{K}_\infty$, we arrive at the QP safety filter

$$\bar{u}_{\mathrm{QP}}=\min\left\{u_0,-(1+x^2)^2\rho^{-1}\left(\max\{0,|x|x\}\right)-x\right\}. \tag{118}$$

□

Next, we give a result on inverse optimal NSSf filter design.

Theorem 10: Consider the control law

$$u=u_0+\bar{u}(x,u_0) \tag{119}$$

$$\bar{u}(x,u_0)=R_2^{-1}\left(L_{g_2}h\right)^{\mathrm{T}}\frac{\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)}{\left|L_{g_2}hR_2^{-1/2}\right|^2} \tag{120}$$

where $h(x)$ is a barrier function candidate, γ_1 and γ_2 are class $\mathcal{K}_\infty$ functions whose derivatives are also class $\mathcal{K}_\infty$ functions, and $R_2(x,u_0)$ is a matrix-valued function such that $R_2(x,u_0)=R_2(x,u_0)^{\mathrm{T}}>0$. If the control law (120) makes the system

$$\mathrm{d}x=f(x)\,\mathrm{d}t+g_1(x)d\bar{w}+g_2(x)u\,\mathrm{d}t \tag{121}$$

satisfy the SBFc with respect to an NSSf-CBF candidate $h(x)$, where $\bar{w}$ is an r-dimensional stochastic process with incremental covariance

$$\bar{\Sigma}\bar{\Sigma}^{\mathrm{T}}=-2g_1^{\mathrm{T}}\frac{\partial^2h}{\partial x^2}g_1\frac{\ell\gamma_1\left(\left|g_1^{\mathrm{T}}\frac{\partial^2h}{\partial x^2}g_1\right|_{\mathcal{F}}\right)}{\left|g_1^{\mathrm{T}}\frac{\partial^2h}{\partial x^2}g_1\right|_{\mathcal{F}}^2} \tag{122}$$

namely, if the condition

$$L_{f+g_2u_0}h-\ell\gamma_1\left(\left|g_1^{\mathrm{T}}\frac{\partial^2h}{\partial x^2}g_1\right|_{\mathcal{F}}\right)+\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)\geq-\alpha(h) \tag{123}$$

is satisfied, then the control law

$$u=u_0+\bar{u}^*(x,u_0) \tag{124}$$

$$\bar{u}^*=\frac{\beta}{2}R_2^{-1}\left(L_{g_2}h\right)^{\mathrm{T}}\frac{(\gamma_2')^{-1}\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)}{\left|L_{g_2}hR_2^{-1/2}\right|},\quad\beta\geq2 \tag{125}$$

maximizes the cost functional

$$J(u-u_0)=\inf_{\Sigma\in\mathcal{D}}\Bigg\{\lim_{t\to\infty}E\left[2\beta h(x(t))\right.$$

$$+\int_0^t \left(l(x,u_0)-\beta^2\gamma_2\left(\frac{2}{\beta}\left|R_2^{1/2}(u-u_0)\right|\right)\right.$$
$$\left.+\beta\lambda\gamma_1\left(\frac{\left|\Sigma\Sigma^{\mathrm{T}}\right|_{\mathcal{F}}}{\lambda}\right)\right)d\tau\Bigg]\Bigg\} \tag{126}$$

where $\lambda\in(0,2]$ and

$$\begin{aligned} l(x,u_0) = & -2\beta\Bigg[L_{f+u_0}h-\ell\gamma_1\left(\left|g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right|_{\mathcal{F}}\right)\\ &+\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)\Bigg]\\ &-\beta(\beta-2)\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)\\ &-\beta(2-\lambda)\ell\gamma_1\left(\left|g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right|_{\mathcal{F}}\right)\\ &\le 2\beta\alpha(h)\,. \end{aligned} \tag{127}$$

Proof: According to Dynkin's formula and by substituting $l(x,u_0)$ into $J(u-u_0)$, we have

$$\begin{aligned} J(u-u_0) = & \inf_{\Sigma\in\mathcal{D}}\Bigg\{\lim_{t\to\infty}E\Bigg[2\beta h(x(t))\\ &+\int_0^t\left(l(x,u_0)-\beta^2\gamma_2\left(\frac{2}{\beta}\left|R_2^{1/2}(u-u_0)\right|\right)\right.\\ &\left.+\beta\lambda\gamma_1\left(\frac{\left|\Sigma\Sigma^{\mathrm{T}}\right|_{\mathcal{F}}}{\lambda}\right)\right)d\tau\Bigg]\Bigg\}\\ = & \inf_{\Sigma\in\mathcal{D}}\Bigg\{\lim_{t\to\infty}E\left[2\beta h(x(0))\right.\\ &+\int_0^t\left(2\beta\mathcal{L}h\,|_{(83)}+l(x,u_0)\right.\\ &-\beta^2\gamma_2\left(\frac{2}{\beta}\left|R_2^{1/2}(u-u_0)\right|\right)\\ &\left.\left.+\beta\lambda\gamma_1\left(\frac{\left|\Sigma\Sigma^{\mathrm{T}}\right|_{\mathcal{F}}}{\lambda}\right)\right)d\tau\right]\Bigg\}\\ = & \inf_{\Sigma\in\mathcal{D}}\Bigg\{2\beta E\left\{V(x(0))\right\}\\ &+\lim_{t\to\infty}E\int_0^t\left[-\beta^2\gamma_2\left(\frac{2}{\beta}\left|R_2^{1/2}(u-u_0)\right|\right)\right.\\ &-\beta^2\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)+2\beta L_{g_2}h(u-u_0)\\ &+\beta\lambda\gamma_1\left(\frac{\left|\Sigma\Sigma^{\mathrm{T}}\right|_{\mathcal{F}}}{\lambda}\right)+\beta\lambda\ell\gamma_1\left(\left|g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right|_{\mathcal{F}}\right)\\ &\left.+\beta\mathrm{Tr}\left\{\Sigma^{\mathrm{T}}g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\Sigma\right\}\right]d\tau\Bigg\}. \end{aligned} \tag{128}$$

Using Lemma 5, we have

$$\begin{aligned} &-2\beta L_{g_2}h(u-u_0)\\ &=\beta^2\left(\frac{2}{\beta}R_2^{1/2}(u-u_0)\right)^{\mathrm{T}}\left(-R_2^{-1/2}\left(L_{g_2}h\right)^{\mathrm{T}}\right)\\ &\le\beta^2\gamma_2\left(\frac{2}{\beta}\left|R_2^{1/2}(u-u_0)\right|\right)+\beta^2\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right) \end{aligned} \tag{129}$$

and

$$\begin{aligned} &\beta\mathrm{Tr}\left\{\Sigma^{\mathrm{T}}g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\Sigma\right\}\\ &=\beta\left(\mathrm{col}\left(\Sigma\Sigma^{\mathrm{T}}\right)\right)^{\mathrm{T}}\left(\mathrm{col}\left(g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right)\right)\\ &\le\beta\lambda\gamma_1\left(\frac{\left|\Sigma\Sigma^{\mathrm{T}}\right|_{\mathcal{F}}}{\lambda}\right)+\beta\lambda\ell\gamma_1\left(\left|g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right|_{\mathcal{F}}\right) \end{aligned} \tag{130}$$

and the equalities hold when (125) and

$$\left(\Sigma\Sigma^{\mathrm{T}}\right)^*=-\lambda(\gamma_1')^{-1}\left(\left|g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right|_{\mathcal{F}}\right)\frac{g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1}{\left|g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right|_{\mathcal{F}}}. \tag{131}$$

So the "worst-case" unknown covariance is given by (131), the minimum of (128) is reached with $u=\bar{u}^*$, and $\min_{u-u_0}J(u-u_0)=2\beta E\{h(x(0))\}$. □

Remark 6: Similar to Remarks 3 and 5, even though not explicit in the statement of Theorem 10, $h(x)$ solves the following family of *HJI* equations parameterized by $\beta\in[2,\infty)$ and $\lambda\in(0,2]$:

$$\begin{aligned} L_{f+u_0}V-\frac{\lambda}{2}\ell\gamma_1\left(\left|g_1^{\mathrm{T}}\frac{\partial^2 h}{\partial x^2}g_1\right|_{\mathcal{F}}\right)+\frac{\beta}{2}\ell\gamma_2\left(\left|L_{g_2}hR_2^{-1/2}\right|\right)\\ +\frac{l(x)}{2\beta}=0. \end{aligned} \tag{132}$$

This equation, which depends only on known quantities, helps explain why we are pursuing a differential game formulation for safe control design, with Σ as a player. □

Remark 7: Similar to Remark 4, we refer to the property

$$\lim_{t\to\infty}E\left\{h(x(t))+\int_0^t\left[\alpha(h(x))+\frac{\lambda}{2}\gamma_1\left(\frac{\left|\Sigma\Sigma^{\mathrm{T}}\right|_{\mathcal{F}}}{\lambda}\right)\right]\mathrm{d}t\right\}\ge 0 \tag{133}$$

as integral noise-to-state safety. □

X. Inverse Optimal Safety Filters for Adaptive Control, System Identification, and Extremum Seeking

Most problems involving estimation of unknown parameters—be it in system identification, adaptive control, or extremum seeking—involve optimization. Safe sets of unknown parameters are not just about producing estimates that are within a set in which the unknown parameter is known to be. There is a more critical reason for keeping the parameters in a "safe set" in adaptive control—the safe set typically contains parameter values that correspond to the system model being controllable or stabilizable. Employing parameter estimates from outside of the safe set in indirect adaptive control results in an attempt of stabilizing an unstabilizable system, the result of which is the controller gains assuming infinite values. Hence, keeping parameters inside a safe set is critical in these domains of control theory.

Keeping parameters inside a safe set is an add-on to parameter estimator design. The typical problem of safety maintenance is simple. The parameter estimator has the form of a vector integrator

$$\dot{\hat{\theta}} = u \tag{134}$$

where $\hat{\theta} \in \mathbb{R}^n$ is the parameter estimate and $u = u_0(\hat{\theta}, \xi, t)$ is the parameter estimator feedback, designed using gradient, least-squares, Lyapunov, passivity, or some other method, and possibly dependent on an additional state ξ, which may contain the states of the plant, an observer, and filters. The safety objective is formulated as keeping $\hat{\theta}$ inside the set $\{h(\hat{\theta}) \geq 0\}$, where h has the usual properties of a CBF.

The conventional safety filter in parameter estimation is *parameter projection*. Parameter projection assigns

$$u = u_0 + \bar{u} \tag{135}$$

where $\bar{u} = \bar{u}_{\rm P}$ and

$$\bar{u}_{\rm P} = \frac{\left(\frac{\partial h}{\partial \hat{\theta}}\right)^T}{\left|\frac{\partial h}{\partial \hat{\theta}}\right|^2} \begin{cases} 0, & \alpha\left(h\left(\hat{\theta}\right)\right) > 0 \text{ or } \frac{\partial h}{\partial \hat{\theta}} u_0 \geq 0 \\ -\frac{\partial h}{\partial \hat{\theta}} u_0, & \alpha\left(h\left(\hat{\theta}\right)\right) = 0 \text{ and } \frac{\partial h}{\partial \hat{\theta}} u_0 < 0 \end{cases} \tag{136}$$

and $\alpha \in \mathcal{K}$ is arbitrary, typically taken as identity.

Let us contrast this with the safety filter obtained using the QP approach where $\bar{u} = \bar{u}_{\rm QP}$ and

$$\bar{u}_{\rm QP} = \frac{\left(\frac{\partial h}{\partial \hat{\theta}}\right)^T}{\left|\frac{\partial h}{\partial \hat{\theta}}\right|^2} \max\left\{0, -\frac{\partial h}{\partial \hat{\theta}} u_0 - \alpha\left(h\left(\hat{\theta}\right)\right)\right\}. \tag{137}$$

Recall Remark 2 on the notational convention of impossibility of division by $\frac{\partial h}{\partial \hat{\theta}} = 0$ in (137).

The similarity between (136) and (137) is striking and not noted before in the literature, as (137) has not seen use in parameter estimation. While (137) is continuous, (136) interferes less with the nominal u_0 (it lets the trajectory come to the boundary and then glide tangentially if u_0 demands an exit from the safe set) but is discontinuous (at the boundary of the safe set). Intuitively, if one were to take $\alpha(r)$ outside of class $\mathcal{K}$, as a mapping such that $\alpha(0) = 0$ but $\alpha(r) = +\infty$ for all $r > 0$, one would obtain (136) from (137). One can approximate the projection operator quite closely by (137) if one takes $\alpha(r) = \frac{1}{\epsilon} r^\epsilon$ for sufficiently small positive ϵ.

Neither (136) nor (137) have optimality properties, but the following result, proven using Corollary 1, holds for (137).

Theorem 11: The update law (134) with

$$u = u_0 + \beta \bar{u}_{\rm QP} \tag{138}$$

and (137), for any $\beta \geq 2$, *minimizes*

$$J(u - u_0) = \lim_{t \to \infty} \left[-2\beta h\left(\hat{\theta}\right) + \int_0^t \left(l\left(\hat{\theta}, u_0\right) + \frac{\left|\frac{\partial h}{\partial \hat{\theta}}\right|^2 |u - u_0|^2}{\max\left\{0, -\frac{\partial h}{\partial \hat{\theta}} u_0 - \alpha\left(h\left(\hat{\theta}\right)\right)\right\}} \right) {\rm d}\tau \right] \tag{139}$$

where

$$l\left(\hat{\theta}, u_0\right) = 2\beta \frac{\partial h}{\partial \hat{\theta}} u_0 + \beta^2 \max\left\{0, -\frac{\partial h}{\partial \hat{\theta}} u_0 - \alpha\left(h\left(\hat{\theta}\right)\right)\right\} \geq -2\beta\alpha\left(h\left(\hat{\theta}\right)\right). \tag{140}$$

The cost functional (139) clearly favors the update law u staying close to the nominal design u_0. The two occurrences of a negative of $h(x)$ in (139), first in the "terminal penalty" before the integral and, second, in the lower bound on the state penalty $l(x, u_0)$ should be understood as measures of "nonsafety" of the parameter estimator. By minimizing these nonsafety measures, the safety filter (138) maximizes the estimator's safety.

To summarize the update law, (134), (138), and (137), we get

$$\dot{\hat{\theta}} = u_0 + \beta \frac{\left(\frac{\partial h}{\partial \hat{\theta}}\right)^T}{\left|\frac{\partial h}{\partial \hat{\theta}}\right|^2} \max\left\{0, -\frac{\partial h}{\partial \hat{\theta}} u_0 - \alpha\left(h\left(\hat{\theta}\right)\right)\right\} \tag{141}$$

for any $\beta \geq 2$. The "soft projection" in [29, E.5 and E.6] is a special case of the update law (141).

XI. Regulation to the Safety Boundary

In some applications, the objective is not to keep the state away from the boundary but to regulate it *to* the boundary, without a safety violation. A very special case of such an objective, for strict-feedback systems, with the safe set being a "half-space" and its boundary a hyperplane, was pursued in [28], under the name of "nonovershooting" control and without employing any safety filters. In addition, in Section IV of that paper, a problem of input-to-state stabilization was tackled, where convergence to the boundary was the goal but the achievement of regulation to a neighborhood of the boundary proportional to the unknown bound on the disturbance was guaranteed.

In general, one can expect that the objective of u_0 be stabilization of the entire boundary, namely, regulation to the boundary (on which there may not even be an equilibrium), rather than to a unique equilibrium on the boundary. We pursue in this section the regulation for general boundaries/CBFs and for general systems

$$\dot{x} = f(x) + g_1(x)d + g_2(x)u\,. \tag{142}$$

We do not employ any Lyapunov functions and we do not consider nominal u_0 whose task is equilibrium stabilization. Instead, the objective of u_0 is regulation of the barrier function $h(x(t))$ to zero, just as the objective of the safety filter $\bar{u}(x, u_0)$ shall be to prevent $h(x(t))$ from approaching zero too fast.

To summarize, we consider a single barrier function $h(x)$ with two contradictory objectives but with two distinct input functions to be designed: $u_0(x)$ tasked with reducing $h(x)$ to zero and $\bar{u}(x, u_0)$ tasked with keeping $h(x)$ away from zero for all finite time. We approach these two simultaneous tasks with the following two functions:

$$\omega_0(x) = L_{f+g_2 u_0(x)} h + |L_{g_1} h|\, \rho_0^{-1}(\max\{0, h(x)\})$$

$$+\alpha_0(h(x)) \tag{143}$$

$$\omega(x,u_0) = L_{f+g_2u_0}h - |L_{g_1}h|\,\rho^{-1}(\max\{0,-h(x)\}) + \alpha(h(x)) \tag{144}$$

where $\rho_0 : [0,+\infty) \to [0, \sup h(\xi))$ and $\rho : [0,+\infty) \to [0, -\inf h(\xi))$ are in class $\mathcal{K}$ and $\alpha_0, \alpha \in \mathcal{K}_h$.

We urge the reader to closely examine the similarities and differences between these two functions and then proceed.

Assumption 4: The feedback law $u_0(x)$ is input-to-output stabilizing for the system (142) from the disturbance d to the CBF $h(x)$ as an output, namely

$$\omega_0(x) \le 0 \quad \forall x \in \mathbb{R}^n\,. \tag{145}$$

Assumption 5: For a given nominal feedback law $u_0(x)$, the function $h(x)$ is an ISSf-CBF for system (142), namely

$$L_{g_2}h(x) = 0 \quad \Rightarrow \quad \omega(x,u_0(x)) \ge 0 \quad \forall x \in \mathbb{R}^n\,. \tag{146}$$

Assumption 6: For all $x \in \mathbb{R}^n$,

$$\alpha(h(x)) \ge \alpha_0(h(x)) + |L_{g_1}h|\left[\rho^{-1}(\max\{0,-h(x)\}) + \rho_0^{-1}(\max\{0,h(x)\})\right]. \tag{147}$$

The meanings of Assumptions 4 and 5 are obvious. Assumption 6 means that the system is such that its control can impart a sufficient margin of safety relative to the nominal control $u_0(x)$ designed to drive the system to the safety boundary. The systems in [28], as well as their rather complicated backstepping designs of $u_0(x)$, satisfy all these assumptions.

A generalization of nonovershooting control is given next.

Theorem 12: Under Assumptions 4–6, there exist functions $\beta \le \beta_0$ of class $\mathcal{KL}_h$ such that

$$u = u_0(x) + (L_{g_2}h)^T \frac{\max\{0, -\omega(x,u_0(x))\}}{|L_{g_2}h|^2} \tag{148}$$

ensures the following for system (142) for all $t \ge 0$:

$$\beta(h(x_0),t) - \rho\left(\sup_{0\le\tau\le t}|d(\tau)|\right) \tag{149}$$

$$\le h(x(t)) \le$$

$$\beta_0(h(x_0),t) + \rho_0\left(\sup_{0\le\tau\le t}|d(\tau)|\right). \tag{150}$$

Proof: Result (149) follows from Theorem 2. To prove (150), we substitute (148) and (144) into (142) and get

$$\begin{aligned}\dot h &= L_{f+g_2u_0}h + L_{g_1}hd + L_{g_2}h\bar u_{\rm QP}\\ &= -\alpha_0(h(x)) + \omega_0 + \max\{0,-\omega\}\\ &\quad - |L_{g_1}h|\rho^{-1}(\max\{0,h(x)\}) + L_{g_1}hd\\ &\le -\alpha_0(h(x)) + \max\{\omega_0, \omega_0-\omega\}\\ &\quad - |L_{g_1}h|\left[\rho_0^{-1}(\max\{0,h(x)\}) - |d|\right].\end{aligned} \tag{151}$$

Since $\omega_0 - \omega = \alpha_0 - \alpha \le 0$ and $\omega_0 \le 0$, we have that $\max\{\omega_0, \omega_0-\omega\} \le 0$ and, hence,

$$\dot h \le -\alpha_0(h(x)) - |L_{g_1}h|\left[\rho_0^{-1}(\max\{0,h(x)\}) - |d|\right]. \tag{152}$$

With an argument as, for example, in the proof given in [26, Th. 2.2], (150) follows. □

For the disturbance-free version of (142), namely, for $g_1 = 0$, Theorem 12 yields the following corollary.

Corollary 3: For a given feedback law $u_0(x)$ for system

$$\dot x = f(x) + g(x)u \tag{153}$$

let there exist functions $\alpha_0 \le \alpha$ in class $\mathcal{K}$ such that, for all $x \in \mathbb{R}^n$, $\omega_0(x) := L_{f+gu_0(x)}h + \alpha_0(h(x)) \le 0$ and $\omega(x,u_0) := L_{f+gu_0}h + \alpha(h(x)) \ge 0$ whenever $L_gh = 0$. Then, there exist functions $\beta \le \beta_0$ of class $\mathcal{KL}$ such that

$$u = u_0(x) + (L_gh)^T \frac{\max\{0,-\omega(x,u_0(x))\}}{|L_gh|^2} \tag{154}$$

ensures the following for all $t \ge 0$:

$$\beta(h(x_0),t) \le h(x(t)) \le \beta_0(h(x_0),t)\,. \tag{155}$$

A similar nonovershooting design can be applied to the stochastic system (83) with unknown covariance (112). Inspired by (116), the functions in (143) and (144) are modified, respectively, to

$$\omega_0(x) = L_{f+g_2u_0(x)}h + \alpha_0(h(x)) + \frac{1}{2}\left|g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\right|_{\mathcal{F}}\rho_0^{-1}(\max\{0,h(x)\}) \tag{156}$$

$$\omega(x,u_0) = L_{f+g_2u_0}h + \alpha(h(x)) - \frac{1}{2}\left|g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\right|_{\mathcal{F}}\rho^{-1}(\max\{0,-h(x)\})\,. \tag{157}$$

Denoting $\underline{\rho}(r) = \min\{\rho_0(r), \rho(r)\}$, it can be proven that feedback (148) guarantees that, for all $x \in \mathbb{R}^n$

$$\left|\Sigma\Sigma^{\rm T}\right|_{\mathcal{F}} \ge \underline{\rho}(|h(x)|) \;\Rightarrow\; -\alpha(h(x)) \le \mathcal{L}h(x) \le -\alpha_0(h(x))\,.$$

Finally, for the covariance $\Sigma(t)\Sigma(t)^{\rm T} \equiv I$, the stochastic nonovershooting-in-the-mean results in [30] are generalized as follows. Consider the stochastic system (83) and, for a given feedback law $u_0(x)$, let there exist functions $\alpha_0 \le \alpha$ in class $\mathcal{K}$ such that, for all $x \in \mathbb{R}^n$, $\omega_0(x) := L_{f+g_2u_0(x)}h + \frac{1}{2}{\rm Tr}\{g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\} + \alpha_0(h(x)) \le 0$ and $\omega(x,u_0) := L_{f+g_2u_0}h + \frac{1}{2}{\rm Tr}\{g_1^{\rm T}\frac{\partial^2 h}{\partial x^2}g_1\} + \alpha(h(x)) \ge 0$ whenever $L_{g_2}h = 0$. Then, (148) ensures that, for all $x \in \mathbb{R}^n$

$$-\alpha(h(x)) \le \mathcal{L}h(x) \le -\alpha_0(h(x))\,. \tag{158}$$

In this section, h played a twofold role of a barrier and Lyapunov function. For a general method for simultaneous (vector) Lyapunov-like functions for the same system, see [24].

XII. Conclusion

For nonlinear systems affine in control, deterministic disturbance, or stochastic disturbance, we introduced the appropriate notions of CBFs, designed safety-ensuring filters, and produced parameterized families of safety filters that have inverse optimality properties. Optimality is always such that the safety filter is rewarded for increasing safety and for keeping the input close to nominal, whereas the disturbance is rewarded for decreasing safety and for not spending high energy.

Theorems 4 and 10, as well as Example 4, show the potential benefit of stepping beyond the confines of the QP/min-norm design.

The theory presented here is of value only if CBFs (as well as ISSf-CBFs and NSSf-CBFs) can be systematically constructed, for given constraints on the state, i.e., for given admissible sets.

Designs of CLFs were, similarly, far from straightforward but the classes of strict-feedback/triangular systems have proven fruitful in [29] and in the subsequent literature. Similar promise resides for CBFs for strict-feedback systems, as illustrated in [28] and in the subsequent articles [10], [20], [35], [51], [52], [54].

It is well known that not only may safety filters interfere with stabilization [41], but the loss of boundedness may be inevitable if enforcing state constraints, as in the case of constraints of a nonminimum phase kind, where trajectories from a portion of the state space that honor the constraint always go to infinity regardless of the controller [42]. This is why we limit our attention to the topics of safety and liveness in this article, without using stability to measure success with these distinct objectives.

Acknowledgment

The author would like to thank Iasson Karafyllis, Xiangru Xu, and Andrew Clark for helpful feedback on a draft of the paper.

Appendix

Appendix Legendre-Fenchel Transform and Young's Inequality

Lemma 4 [27, Lemma A.1]: If γ and its derivative γ' are class $\mathcal{K}_\infty$, then the Legendre–Fenchel transform satisfies the following properties:

$$1)\ \ell\gamma(r) = r(\gamma')^{-1}(r) - \gamma\left((\gamma')^{-1}(r)\right) = \int_0^r (\gamma')^{-1}(s)ds \tag{159}$$

$$2)\ \ell\ell\gamma = \gamma \tag{160}$$

$$3)\ \ell\gamma \text{ is a class } \mathcal{K}_\infty \text{ function} \tag{161}$$

$$4)\ \ell\gamma(\gamma'(r)) = r\gamma'(r) - \gamma(r)\,. \tag{162}$$

Lemma 5 (Young's inequality [19, Th. 156]): For any $x, y \in \mathbb{R}^n$, and for any $\gamma \in \mathcal{K}_\infty$ whose derivative is also in $\mathcal{K}_\infty$,

$$x^{\mathrm{T}} y \le \gamma(|x|) + \ell\gamma(|y|) \tag{163}$$

and the equality is achieved if and only if

$$y = \gamma'(|x|)\frac{x}{|x|}\,, \text{ that is, for } x = (\gamma')^{-1}(|y|)\frac{y}{|y|}\,. \tag{164}$$

References

[1] I. Abel, M. Janković, and M. Krstić, "Constrained control of input delayed systems with partially compensated input delays," in *Proc. ASME Dyn. Syst. Controls Conf.*, 2020.

[2] H. Almubarak, N. Sadegh, and E. A. Theodorou, "Safety embedded control of nonlinear systems via barrier states," *IEEE Contr. Syst. Lett.*, vol. 6, pp. 1328–1333, 2022.

[3] H. Almubarak, E. A. Theodorou, and N. Sadegh, "HJB based optimal safe control using control barrier functions," in *Proc. IEEE 60th Conf. Decis. Control*, 2021, pp. 6829–6834.

[4] A. D. Ames, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs with application to adaptive cruise control," in *Proc. IEEE Conf. Decis. Control*, 2014, pp. 6271–6278.

[5] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Trans. Autom. Control*, vol. 62, no. 8, pp. 3861–3876, Aug. 2017.

[6] D. Angeli and E. D. Sontag, "Forward completeness, unboundedness observability, and their Lyapunov characterizations," *Syst. Control Lett.*, vol. 38, no. 4, pp. 209–217, 1999.

[7] T. Başar and P. Bernhard, *H^∞-Optimal Control and Related Minimax Design Problems: A. Dynamic Game Approach*. Basel, Switzerland: Birkhauser, 1998.

[8] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed. Philadelphia, PA, USA: SIAM, 1998.

[9] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin, "Hamilton-Jacobi reachability: A brief overview and recent advances," in *Proc. IEEE 56th Annu. Conf. Decis. Control*, 2017, pp. 2242–2253.

[10] J. Breeden and D. Panagou, "High relative degree control barrier functions under input constraints," in *Proc. IEEE 60th Conf. Decis. Control*, 2021, pp. 6119–6124.

[11] Y. Chen, M. Ahmadi, and A. D. Ames, "Optimal safe controller synthesis: A density function approach," in *Proc. Amer. Control Conf.*, 2020, pp. 5407–5412.

[12] J. J. Choi, D. Lee, K. Sreenath, C. J. Tomlin, and S. L. Herbert, "Robust control barrier-value functions for safety-critical control," in *Proc. IEEE 60th Conf. Decis. Control*, 2021, pp. 6814–6821.

[13] A. Clark, "Control barrier functions for stochastic systems," *Automatica*, vol. 130, 2021, Art. no. 109688.

[14] M. H. Cohen and C. Belta, "Approximate optimal control for safety-critical systems with control barrier functions," in *Proc. IEEE 59th Conf. Decis. Control*, 2020, pp. 2062–2067.

[15] H. Deng, M. Krstic, and R. Williams, "Stabilization of stochastic nonlinear systems driven by noise of unknown covariance," *IEEE Trans. Autom. Control*, vol. 46, no. 8, pp. 1237–1253, 2001.

[16] H. Deng and M. Krstic, "Stochastic nonlinear stabilization–II: Inverse optimality," *Syst. Control Lett.*, vol. 32, pp. 151–159, 1997.

[17] R. A. Freeman and P. V. Kokotovic, "Inverse optimality in robust stabilization," *SIAM J. Control Optim.*, vol. 34, pp. 1365–1391, 1996.

[18] P. Glotfelter, J. Cortés, and M. Egerstedt, "Nonsmooth barrier functions with applications to multi-robot systems," *IEEE Contr. Syst. Lett.*, vol. 1, no. 2, pp. 310–315, Oct. 2017.

[19] G. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*. Cambridge, U.K.: Cambridge Univ. Press, 1989.

[20] S.-C. Hsu, X. Xu, and A. D. Ames, "Control barrier function based quadratic programs with application to bipedal robotic walking," in *Proc. Amer. Control Conf.*, 2015, pp. 4542–4548.

[21] H. Ito and R. A. Freeman, "Uniting local and global controllers for uncertain nonlinear systems: Beyond global inverse optimality," *Syst. Control Lett.*, vol. 45, no. 1, pp. 59–79, 2002.

[22] M. Janković, "Control barrier functions for constrained control of linear systems with input delay," in *Proc. Amer. Control Conf.*, 2018, pp. 3316–3321.

[23] M. Jankovic, "Robust control barrier functions for constrained stabilization of nonlinear systems," *Automatica*, vol. 96, pp. 359–367, 2018.

[24] I. Karafyllis and Z.-P. Jiang, "Global stabilization of nonlinear systems based on vector control Lyapunov functions," *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2550–2562, Oct. 2013.

[25] S. Kolathaya and A. D. Ames, "Input-to-state safety with control barrier functions," *IEEE Contr. Syst. Lett.*, vol. 3, pp. 108–113, Jan. 2019.

[26] M. Krstić and H. Deng, *Stabilization of Nonlinear Uncertain Systems*. Berlin, Germany: Springer, 2000.

[27] M. Krstic and Z.-H. Li, "Inverse optimal design of input-to-state stabilizing nonlinear controllers," *IEEE Trans. Autom. Control*, vol. 43, no. 3, pp. 336–350, Mar. 1998.

[28] M. Krstic and M. Bement, "Nonovershooting control of strict-feedback nonlinear systems," *IEEE Trans. Autom. Control*, vol. 51, no. 12, pp. 1938–1943, Dec. 2006.

[29] M. Krstic, P. V. Kokotovic, and I. Kanellakopoulos, *Nonlinear and Adaptive Control Design*. Hoboken, NJ, USA: Wiley, 1995.

[30] W. Li and M. Krstic, "Mean-nonovershooting control of stochastic nonlinear systems," *IEEE Trans. Autom. Control*, vol. 66, no. 12, pp. 5756–5771, Dec. 2021.

[31] Z.-H. Li and M. Krstic, "Optimal design of adaptive tracking controllers for non-linear systems," *Automatica*, vol. 33, pp. 1459–1473, 1997.

[32] Z. Lyu, X. Xu, and Y. Hong, "Small-gain theorem for safety verification of interconnected systems," *Automatica*, vol. 139, 2022, Art. no. 110178. 2020.

[33] M. Maghenem and R. G. Sanfelice, "Characterization of safety and conditional invariance for nonlinear systems," in *Proc. Amer. Control Conf.*, 2019, pp. 5039–5044.

[34] T. G. Molnár, A. W. Singletary, G. Orosz, and A. D. Ames, "Safety-critical control of compartmental epidemiological models with measurement delays," *IEEE Contr. Syst. Lett.*, vol. 5, no. 5, pp. 1537–1542, Nov. 2021.

[35] Q. Nguyen and K. Sreenath, "Exponential control barrier functions for enforcing high relative-degree safety-critical constraints," in *Proc. Amer. Control Conf.*, 2016, pp. 322–328.
[36] Z. Pan, K. Ezal, A. Krener, and P. Kokotovic, "Backstepping design with local optimality matching," *IEEE Trans. Autom. Control*, vol. 46, no. 7, pp. 1014–1027, Jul. 2001.
[37] S. Prajna and A. Jadbabaie, "Methods for safety verification of time-delay systems," in *Proc. IEEE 44th Conf. Decis. Control*, 2005, pp. 4348–4353.
[38] S. Prajna, A. Jadbabaie, and G. J. Pappas, "A framework for worst-case and stochastic safety verification using barrier certificates," *IEEE Trans. Autom. Control*, vol. 52, no. 8, pp. 1415–1428, Aug. 2007.
[39] Y. Rahman, M. Jankovic, and M. Santillo, "Driver intent prediction with barrier functions," in *Proc. Amer. Control Conf.*, 2021, pp. 224–230.
[40] J. Rawlings, D. Mayne, and M. Diehl, *Model Predictive Control: Theory, Computation, and Design*. Madison, WI, USA: Nob Hill Publishing, 2017.
[41] M. F. Reis, A. P. Aguiar, and P. Tabuada, "Control barrier function-based quadratic programs introduce undesirable asymptotically stable equilibria," *IEEE Control Syst. Lett.*, vol. 5, no. 2, pp. 731–736, Apr. 2021.
[42] A. Saberi, J. Han, and A. A. Stoorvogel, "Constrained stabilization problems for linear plants," *Automatica*, vol. 38, no. 4, pp. 639–654, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109801002485
[43] M. Santillo and M. Jankovic, "Collision free navigation with interacting, non-communicating obstacles," in *Proc. Amer. Control Conf.*, 2021, pp. 1637–1643.
[44] C. Santoyo, M. Dutreix, and S. Coogan, "A barrier function approach to finite-time stochastic system verification and control," *Automatica*, vol. 125, 2021, Art. no. 109439.
[45] R. Sepulchre, M. Janković, and P. Kokotović, *Constructive Nonlinear Control*. Berlin, Germany: Springer, 1997.
[46] E. Sontag, "Smooth stabilization implies coprime factorization," *IEEE Trans. Autom. Control*, vol. 34, no. 4, pp. 435–443, Apr. 1989.
[47] E. D. Sontag, "A universal construction of Artstein's theorem on nonlinear stabilization," *Syst. Control Lett.*, vol. 13, pp. 117–123, 1989.
[48] K. P. Tee, S. S. Ge, and E. H. Tay, "Barrier lyapunov functions for the control of output-constrained nonlinear systems," *Automatica*, vol. 45, no. 4, pp. 918–927, Apr. 2009.
[49] L. Wang, A. D. Ames, and M. Egerstedt, "Safety barrier certificates for collisions-free multirobot systems," *IEEE Trans. Robot.*, vol. 33, no. 3, pp. 661–674, Jun. 2017.
[50] P. Wieland and F. Allgöwer, "Constructive safety using control barrier functions," *IFAC Proc. Volumes*, 2007, vol. 40, no. 12, pp. 462–467.
[51] G. Wu and K. Sreenath, "Safety-critical and constrained geometric control synthesis using control Lyapunov and control barrier functions for systems evolving on manifolds," in *Proc. Amer. Control Conf.*, 2015, pp. 2038–2044.
[52] W. Xiao and C. Belta, "Control barrier functions for systems with high relative degree," in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 474–479.
[53] X. Xu, P. Tabuada, J. W. Grizzle, and A. D. Ames, "Robustness of control barrier functions for safety critical control," *IFAC-PapersOnLine*, vol. 48, no. 27, pp. 54–61, 2015.
[54] X. Xu, "Constrained control of input-output linearizable systems using control sharing barrier functions," *Automatica*, vol. 87, pp. 195–201, 2018.
[55] X. Xu, J. W. Grizzle, P. Tabuada, and A. D. Ames, "Correctness guarantees for the composition of lane keeping and adaptive cruise control," *IEEE Trans. Automat. Sci. Eng.*, vol. 15, no. 3, pp. 1216–1229, Jul. 2018.
[56] H. Yin, A. Packard, M. Arcak, and P. Seiler, "Reachability analysis using dissipation inequalities for uncertain nonlinear systems," *Syst. Control Lett.*, vol. 142, 2020, Art. no. 104736.
[57] H. Yin, P. Seiler, and M. Arcak, "Backward reachability using integral quadratic constraints for uncertain nonlinear systems," *IEEE Contr. Syst. Lett.*, vol. 5, no. 2, pp. 707–712, Apr. 2021.

Miroslav Krstic (Fellow, IEEE) is currently a Distinguished Professor and Senior Associate Vice Chancellor for Research with UC San Diego, La Jolla, CA, USA.

Prof. Krstic is a Fellow of IFAC, ASME, SIAM, AAAS, IET (U.K.), and AIAA (Associate Fellow), as well as a Foreign Member of the Serbian Academy of Sciences. He was the recipient of the Bode Lecture Prize, Bellman Award, SIAM Reid Prize, ASME Oldenburger Medal, Nyquist Lecture Prize, Paynter Award, Ragazzini Award, IFAC Ruth Curtain Distributed Parameter Systems Award, IFAC Nonlinear Control Systems Award, Chestnut Award, A.V. Balakrishnan Award for the Mathematics of Systems, CSS Distinguished Member Award, the PECASE, NSF Career, ONR YI Awards, the Schuck (96 and 19) and Axelby award, and the first UCSD Research Award given to an engineer.