Systems & Control Letters 161 (2022) 105138

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Remarks on input to state stability of perturbed gradient flows, motivated by model-free feedback control learning

Eduardo D. Sontag *

Bioengineering and Electrical and Computer Engineering Departments, Northeastern University, Boston, MA, United States of America
Affiliate member of Mathematics and Chemical Engineering Departments, Northeastern University, Boston, MA, United States of America
Laboratory of Systems Pharmacology, Program in Therapeutic Science, Harvard Medical School, Boston, MA, United States of America

ARTICLE INFO

Article history:
Received 26 August 2021
Received in revised form 29 December 2021
Accepted 6 January 2022
Available online xxxx

Keywords:
Input to state stability
Data-driven control
Gradient systems
Steepest descent
Model-free control

ABSTRACT

Recent work on data-driven control and reinforcement learning has renewed interest in a relative old field in control theory: model-free optimal control approaches which work directly with a cost function and do not rely upon perfect knowledge of a system model. Instead, an "oracle" returns an estimate of the cost associated to, for example, a proposed linear feedback law to solve a linear–quadratic regulator problem. This estimate, and an estimate of the gradient of the cost, might be obtained by performing experiments on the physical system being controlled. This motivates in turn the analysis of steepest descent algorithms and their associated gradient differential equations. This note studies the effect of errors in the estimation of the gradient, framed in the language of input to state stability, where the input represents a perturbation from the true gradient. Since one needs to study systems evolving on proper open subsets of Euclidean space, a self-contained review of input to state stability definitions and theorems for systems that evolve on such sets is included. The results are then applied to the study of noisy gradient systems, as well as the associated steepest descent algorithms.
© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Suppose that a function $V : \mathbb{X} \to \mathbb{R}$, defined on an open subset $\mathbb{X}$ of $\mathbb{R}^n$, achieves a global minimum at a compact set $\mathcal{A} \subset \mathbb{X}$, and that its gradient does not vanish except on $\mathcal{A}$. Under appropriate technical conditions, the solutions of the gradient flow $\dot{x} = -\eta \nabla V(x)^T$ (where $\eta > 0$ is a "learning rate") will globally, and even exponentially, converge to $\mathcal{A}$ as $t \to \infty$.

In many data-driven applications, the gradient can be well-estimated numerically. The combination of direct gradient estimation and gradient descent has generated strong recent interest in control theory, and specifically in Reinforcement Learning (RL) model-free control. In order to theoretically better understand the problem, several authors have studied an archetypical control problem, the infinite-horizon Linear Quadratic Regulator (LQR) problem. Since the pioneering work of Kalman in the early 1960s, it has been known that the solution of the LQR problem can be obtained explicitly via a Riccati equation, and many computational packages do so very efficiently. Nonetheless, if the system being controlled is imperfectly known, the function to be optimized is not known except through "queries" involving sampling and experimentation, and in that context direct methods might be of interest. In any event, however, working on a well-understood problem like LQR serves to understand properties of model-free approaches.

It turns out that when the LQR problem is formulated as an optimization over a set of stabilizing feedback matrices, the loss function, while not convex, satisfies strong convergence guarantees [1]. The trick is to employ a reparametrization for the LQR problem that allows solving an associated strongly convex problem. We refer the reader to [1] for details. Note that, in the LQR problem as just described, the open set $\mathbb{X}$ is a set of matrices. Restricting the optimization dynamics to this open set is essential for the approach to work.

In this note, we study a perturbed gradient system (superscript T indicates transpose):

$$\dot{x}(t) = -\eta \nabla V(x(t))^T + B(x(t))u(t). \tag{1}$$

The additive term represents disturbances. For example, if $B(x)$ is the constant matrix with rows $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$, …, $(0, 0, \ldots, 1)$ then we have independent disturbances u_i acting on each coordinate. Without the additive term, this is a standard gradient descent flow. For generality, we allow state-dependent perturbations (non-constant B).

The "disturbance" inputs might represent errors that arise from numerically approximating the gradient from data through two-point estimates as in [1], or from measurement noise. The paper [2] interprets the discretization error when solving ODE's as a perturbation, and relates asymptotic stability for dynamical

* Corresponding author.
E-mail address: sontag@sontaglab.org.

https://doi.org/10.1016/j.sysconle.2022.105138
0167-6911/© 2022 Elsevier B.V. All rights reserved.

systems to families of approximations, specifically applying this to numerical one step schemes for ordinary differential equations.

To quantify the effect of disturbances, we will use the notion of input to state stability (ISS), introduced in [3] (see expositions in [4–7]). We will prove (under technical assumptions on V and its gradient, mainly that they blow up at the boundary of $\mathbb{X}$, so that trajectories cannot escape the constraint set; assumptions which hold in the motivating example from [1]) that the disturbed gradient system is ISS. This implies that if the disturbances or errors are bounded, small, "eventually" small, or convergent, the solutions of the system will inherit the same properties, with well-controlled transient behavior.

The natural setting is that of differential equations that evolve in a nontrivial open subset $\mathbb{X}$ of $\mathbb{R}^n$. An example is a gradient system that uses a loss function associated to a feedback matrix K that is required to stabilize a given linear system $\dot{x} = Ax + Bu$, in the sense that $A - BK$ is a Hurwitz matrix (i.e., it has all its eigenvalues with negative real parts). We can view matrices of size $p \times q$ as elements of $\mathbb{R}^n$, $n = pq$. Since eigenvalues depend continuously on matrix entries (a standard fact, proved for example in the linear algebra appendix in [5]), the set $\mathbb{X} \subset \mathbb{R}^n$ of stabilizing matrices (for a fixed system defined by A and B) is open.

The precise statement of the ISS result requires introducing appropriate notions of stability and "size" of elements in open subsets. We consider such notions here; that material that we discuss on ISS on open subsets is not new. However, these results are not easy to find clearly stated in the literature, and they are of independent interest beyond the study of gradient systems, so we provide proofs of several facts about them for more general systems with inputs. Since it represents no additional complication, we consider stability with respect to compact sets as opposed to single equilibria.

In addition to studying the gradient system, we study the performance of steepest descent, the discrete process in which a line search is performed, iteratively minimizing a cost function in the direction of the gradient. Given a continuously differentiable function $V : \mathbb{X} \to \mathbb{R}$ to be minimized on an open subset $\mathbb{X} \subseteq \mathbb{R}^n$, the steepest descent algorithm consists of the following procedure: given any initial state x^0, one performs a line search in the negative gradient direction so as to minimize $V(x^0 - \lambda \nabla V(x^0)^T)$ over $\lambda \geq 0$; the minimal point then defines a new point x^1, and one then iterates. Observe that this search only makes sense on a maximal interval such that the line segment $\{x^0 - \mu \nabla V(x^0)^T, \mu \in [0, \lambda]\}$ is included in $\mathbb{X}$ (so that one may evaluate V for increasing λ). When the gradient is imperfectly evaluated, the picture is further complicated by the fact that one in fact moves in a direction $x^0 - \lambda[\nabla V(x^0)^T + B(x)u]$, for some unknown additive "noise" input vector u (we include $B(x)$ to allow a state-dependence of the input). This gives an iteration that we write as $x^+ = x - \lambda \left[\nabla V(x)^T + B(x)u\right]$.

It is in principle possible that even for a very small step one cannot diminish the cost at all, and moreover one might even exit the set X altogether for an input of large magnitude. A trivial example of this is provided by $\mathbb{X} = (-1, 1)$, $B(x) = 1$, and $V(x) = x^2/2$. The perturbed steepest descent procedure attempts to move to $x - \lambda(x + u)$. If we take any $x > 0$ and any $u < -x$ then for any step size $\lambda > 0$ the cost increases, which means that the steepest descent procedure will be "stuck" at x. Moreover, for large λ the expression $x - \lambda(x + u)$ gives a result outside $\mathbb{X}$. Of course, this can be fixed if the magnitude of the input u is "not too large" compared to the state x. Indeed, we will show that, under reasonable technical assumptions, the steepest descent procedure is input to state stable as a discrete-time system with respect to disturbances.

2. Size functions on open subsets

We start by introducing a notion of "size" that is well-suited to quantifying global convergence to a given compact set, and which in particular acts as a barrier function preventing escape from $\mathbb{X}$.

Definition 2.1. Let $\mathbb{X}$ be an open subset of $\mathbb{R}^n$ and let $\mathcal{A} \subset \mathbb{X}$ be a compact subset. We will say that

$$\omega : \mathbb{X} \to \mathbb{R}$$

is a *size function for* $(\mathbb{X}, \mathcal{A})$ if ω is:

1. continuous,
2. positive definite with respect to $\mathcal{A}$, that is, $\omega(\mathcal{A}) = 0$ and $\omega(x) > 0$ for all $x \in \mathbb{X}$, $x \notin \mathcal{A}$, and
3. proper, that is, for every real number $r \geq 0$, the sublevel set $S_r := \{x \mid \omega(x) \leq r\}$ is a compact subset of $\mathbb{X}$.

Remark 2.2. Observe that, since $\mathbb{X}$ is an open set, asking that S_r is compact in the induced topology of $\mathbb{X}$ is equivalent to asking that S_r is compact as a subset of $\mathbb{R}^n$. □

Remark 2.3. Let us denote by $|x|$ the standard Euclidean norm in $\mathbb{R}^n$ (any other norm could be used as well). When $\mathbb{X} = \mathbb{R}^n$, a natural choice of size is $\omega(x) = |x|_{\mathcal{A}} = \min_{a \in \mathcal{A}} |x - a|$, the distance to the set $\mathcal{A}$. The notion that we introduce here is based on the beautiful paper of Kurzweil [8] (see also [9]), which studied Lyapunov stability theory on open sets, and is a particular case of "measures" in the sense of Lakshmikantham and coauthors (see e.g. [10]). In [11–13], the concept is called a "proper indicator function" (but we prefer not to use that term, since "indicator function" is typically used for the characteristic function of a set). Another point worth mentioning is that the definition of size function and many of the results can equally well be formulated on a general differentiable manifold $\mathbb{X}$. In that sense, the setup in this note is closely related to the work in [14], in which a variant of ISS for systems evolving in manifolds was considered. □

The following elementary exercise in real analysis provides an intuitive characterization of size functions. We denote by $\partial\mathbb{X}$ the boundary of the set $\mathbb{X}$ (which is empty if and only if $\mathbb{X} = \mathbb{R}^n$).

Lemma 2.4. *The following two statements are equivalent for any function $\omega : \mathbb{X} \to \mathbb{R}$:*

(a) ω is a size function for $(\mathbb{X}, \mathcal{A})$
(b) ω is continuous, positive definite with respect to $\mathcal{A}$, and the following property holds for every sequence $\{x_k \in \mathbb{X}, k \geq 1\}$:

$$\text{if either } x_k \to \partial\mathbb{X} \text{ or } |x| \to \infty, \text{ necessarily } \omega(x_k) \to \infty. \tag{2}$$

Proof. We must show that property (2) is equivalent to compactness of every sublevel set S_r.

Suppose that property (2) is true, and pick any $r \geq 0$. By Remark 2.2, we need to prove that S_r is closed and bounded as a subset of $\mathbb{R}^n$.

We first prove that S_r is closed. Suppose that a sequence $\{x_k\}$ in S_r is such that $x_k \to x \in \mathbb{R}^n$ as $k \to \infty$. We must show that $x \in S_r$. There are two cases to consider: $x \notin \mathbb{X}$ and $x \in \mathbb{X}$. In the first case, being the limit of elements in $\mathbb{X}$, necessarily $x \in \partial\mathbb{X}$. Thus $x_k \to \partial\mathbb{X}$ and, by the assumed property, $\omega(x_k) \to \infty$, contradicting the fact that the sequence $\{\omega(x_k)\}$ is bounded (by r), So this case cannot hold. Thus $x \in \mathbb{X}$, so that $x \in S_r$ because S_r is closed in the relative topology of $\mathbb{X}$. (More explicitly: by continuity of ω, we have that $\omega(x_k) \to \omega(x)$, and hence $\omega(x) \leq r$, so $x \in S_r$.).

Next we prove that S_r is bounded. Suppose by way of contradiction that there is a sequence $\{x_k\}$ in S_r is such that $|x_k| \to \infty$

as $k \to \infty$. Again using the assumed property, it follows that $\omega(x_k) \to \infty$, contradicting that all $\omega(x_k) \le r$. Thus S_r is bounded.

Conversely, suppose that S_r is compact for every $r \ge 0$ and consider a sequence $\{x_k \in \mathbb{X}, k \ge 1\}$. Suppose first that $|x_k| \to \infty$. We need to prove that, for every $r > 0$, there is an integer K so that $k > K \Rightarrow w(x_k) > r$. Suppose that this is not true, i.e., there are some r and a subsequence $k_j \to \infty$ so that $x_{k_j} \in S_r$ for all k_j. Replacing $\{x_k\}$ by this subsequence, we can then assume that $x_k \in S_r$ for all k, and still $|x_k| \to \infty$. Since S_r is compact, there is a convergent subsequence with its limit $x \in S_r$. This contradicts that $|x_k| \to \infty$. Similarly, suppose that $x_k \to \partial\mathbb{X}$. By contradiction, assume again that there are some r and a subsequence $k_j \to \infty$ so that $x_{k_j} \in S_r$ for all k_j. Replacing x_k by the subsequence, we can then assume that $x_k \in S_r$ for all k, and still $x_k \to \partial\mathbb{X}$. By compactness, we can assume, taking a subsequence, that $x_k \to x \in S_r \subseteq \mathbb{X}$ for some x. However, since $x_k \to \partial\mathbb{X}$ (because we have subsequences of a sequence converging to the boundary), this implies that $x \in \partial\mathbb{X}$. We have a contradiction, because $\mathbb{X}$ and $\partial\mathbb{X}$ are disjoint subsets of $\mathbb{R}^n$. This completes the proof. ∎

Given any open set $\mathbb{X} \subseteq \mathbb{R}^n$ and any compact $\mathcal{A} \subset \mathbb{X}$, there are many possible size functions for $(\mathbb{X}, \mathcal{A})$. As we remarked earlier, $|x|_{\mathcal{A}}$ works when $\mathbb{X} = \mathbb{R}^n$. In general, we may use, for example:

$$\omega(x) \;=\; \max\left\{|x|_{\mathcal{A}}\,,\; \frac{1}{\operatorname{dist}(x,\partial\mathbb{X})} - \frac{a}{\operatorname{dist}(\mathcal{A},\partial\mathbb{X})}\right\}$$

for any $a \ge 1$. The case $a = 2$ of this formula was given in [8], and with that choice one has that $\omega(x) = |x|_{\mathcal{A}}$ for all x near $\mathcal{A}$.

2.1. Comparing size functions

In the same manner that any two norms on a finite dimensional space are equivalent, there is a notion of equivalence of size functions.

We denote by $\mathbb{R}_{\ge 0}$ the set of nonnegative real numbers.

Recall that $\mathcal{K}$ is the set of functions $\alpha : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ that are continuous, strictly increasing, and satisfy $\alpha(0) = 0$, and $\mathcal{K}_\infty \subset \mathcal{K}$ is the subset of unbounded functions, that is, $\alpha(r) \to \infty$ as $r \to \infty$. The set $\mathcal{K}$ is closed under sums, products, and compositions, as is the set $\mathcal{K}_\infty$. Moreover, functions in $\mathcal{K}_\infty$ are invertible, and $\alpha^{-1} \in \mathcal{K}_\infty$, so $\mathcal{K}_\infty$ is a group under composition (with identity element the map $\alpha(r) = r$). If $\alpha \in \mathcal{K}$, one also says that "α is of class $\mathcal{K}$" and similarly for $\mathcal{K}_\infty$. These classes of functions have played a central role in dynamical systems since at least the textbook by Hahn [15], and were key in the development of input to state stability notions in [3]. They have many other useful properties, for example the weak subadditivity property $\alpha(r+s) \le \alpha(2r) + \alpha(2s)$; see for instance [4,5]. They allow us to relate size functions. Observe that if ω is a size function and $\alpha \in \mathcal{K}_\infty$, then $\alpha \circ \omega$ is also a size function.

Lemma 2.5. *Suppose that ω is a size function for $(\mathbb{X}, \mathcal{A})$. Then for each $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$\omega(x) < \delta \;\Rightarrow\; |x|_{\mathcal{A}} < \varepsilon.$$

Proof. Since ω is continuous and $\omega(\mathcal{A}) = 0$, there exists $\varepsilon_0 > 0$ such that $|x|_{\mathcal{A}} < \varepsilon_0 \Rightarrow \omega(x) < 1$ and $\mathcal{B}_{\varepsilon_0}(\mathcal{A})$, the closed ball of radius ε_0 around $\mathcal{A}$, is included in $\mathbb{X}$. It follows that for any $x \in \mathbb{X}$ with $|x|_{\mathcal{A}} \le \varepsilon_0$, $\omega(x) \le 1$. Now pick any $\varepsilon > 0$. We let $\bar\varepsilon := \min\{\varepsilon, \varepsilon_0\}$. Consider the set

$$C \;:=\; \{x \mid |x|_{\mathcal{A}} \ge \bar\varepsilon \text{ and } \omega(x) \le 1\}.$$

The set is nonempty: pick any $x \in \mathbb{X}$ with $|x|_{\mathcal{A}} = \varepsilon_0$; then $\omega(x) \le 1$ and also $|x|_{\mathcal{A}} = \varepsilon_0 \ge \bar\varepsilon$. The set C is compact, because it is the intersection of a closed set and a compact set. Also, $\omega(x)$ is nonzero in this set, because ω is positive definite. Therefore there is a positive minimum of w on the set C; we pick δ as this minimum, and thus $x \in C \Rightarrow \omega(x) \ge \delta$. Without loss of generality, we will assume $\delta < 1$ (otherwise, make δ smaller). Now assume that $\omega(x) < \delta$. This means that x is not in C, so either $\omega(x) > 1$ or $|x|_{\mathcal{A}} < \bar\varepsilon$. However, $\omega(x) > 1$ cannot happen, because $\omega(x) < \delta < 1$. Therefore, $|x|_{\mathcal{A}} < \bar\varepsilon \le \varepsilon$, as wanted. ∎

Proposition 2.6. *Suppose that ω_1 and ω_2 are two size functions for $(\mathbb{X}, \mathcal{A})$. Then, there is some $\alpha \in \mathcal{K}_\infty$ such that*

$$\omega_1(x) \;\le\; \alpha(w_2(x)) \text{ for all } x \in \mathbb{X}\,. \tag{3}$$

Proof. Define

$$\widetilde\alpha(r) \;:=\; \max_{\{x \mid \omega_2(x) \le r\}} \omega_1(x).$$

Since the set $\{x \mid \omega_2(x) \le r\}$ is compact, this maximum is well-defined. Note that the inequality (3) holds. Indeed, given any $x \in \mathbb{X}$, let $r := \omega_2(x)$; then $\omega_1(x) \le \widetilde\alpha(r) = \widetilde\alpha(\omega_2(x))$, because x belongs to the set over which we are maximizing. Moreover, $\widetilde\alpha$ is nondecreasing (since as r is larger, one takes a maximum over a larger set). Also, $\widetilde\alpha(0) = 0$ by positive definiteness of ω_1 and ω_2. We prove next that $\widetilde\alpha$ is continuous at 0.

Fix any $\varepsilon > 0$. We want to find a $\delta > 0$ so that $r < \delta \Rightarrow \widetilde\alpha(r) < \varepsilon$. From the definition of $\widetilde\alpha$, it is enough to find a δ such that, for each $r < \delta$:

$$\omega_2(x) \le r \;\Rightarrow\; \omega_1(x) < \varepsilon/2.$$

Since ω_1 is continuous and $\omega_1(\mathcal{A}) = 0$, there is a $\delta_1 > 0$ such that

$$|x|_{\mathcal{A}} < \delta_1 \;\Rightarrow\; \omega_1(x) < \varepsilon/2.$$

By Lemma 2.5 applied to ω_2 and $\varepsilon = \delta_1$, there is a $\delta > 0$ such that

$$\omega_2(x) < \delta \;\Rightarrow\; |x|_{\mathcal{A}} < \delta_1.$$

We conclude that:

$$\omega_2(x) < \delta \;\Rightarrow\; \omega_1(x) < \varepsilon/2.$$

Now assume $r < \delta$. For any x such that $\omega_2(x) \le r$, also $\omega_2(x) < \delta$, and hence $\omega_1(x) < \varepsilon/2$, as wanted.

So far we have a nondecreasing $\widetilde\alpha : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ which satisfies $\widetilde\alpha(0) = 0$ and is continuous at zero. Such a function can be majorized by a class $\mathcal{K}_\infty$ function α, i.e. $\widetilde\alpha(r) \le \alpha(r)$ for all r, which together with $\omega_1(x) \le \widetilde\alpha(\omega_2(x))$ implies the estimate (3). The construction of α is a standard exercise. First majorize $\widetilde\alpha$ by a nondecreasing continuous function. For example, pick a doubly infinite sequence of nonnegative numbers r_k, $k \in \mathbb{Z}$ so that $r_k \to 0$ monotonically as $k \to -\infty$ and $r_k \to \infty$ monotonically as $k \to +\infty$ and let α interpolate linearly the values $(r_k, \widetilde\alpha(r_{k+1}))$ (recall that $\widetilde\alpha$ is nondecreasing, so that the interpolation function is nondecreasing, and it clearly majorizes $\widetilde\alpha$). This gives an $\alpha \in \mathcal{K}$. Finally, add any $\mathcal{K}_\infty$ function to obtain an α of class $\mathcal{K}_\infty$. ∎

Corollary 2.7. *Suppose that ω_1 is a size function for $(\mathbb{X}, \mathcal{A})$. Let $\omega_2 : \mathbb{X} \to \mathbb{R}^n$ be a continuous function. Then the following properties are equivalent:*

(a) ω_2 is a size function for $(\mathbb{X}, \mathcal{A})$;
(b) there exist functions $\alpha_1, \alpha_2 \in \mathcal{K}_\infty$ such that

$$\alpha_1(\omega_1(x)) \;\le\; \omega_2(x) \;\le\; \alpha_2(w_1(x)) \quad \text{for all } x \in \mathbb{X}\,. \tag{4}$$

Proof. Suppose that ω_2 is a size function for $(\mathbb{X}, \mathcal{A})$. By Proposition 2.6, there is an $\alpha \in \mathcal{K}_\infty$ such that $\omega_1(x) \le \alpha(w_2(x))$ for all x. Thus $\alpha_1(\omega_1(x)) \le w_2(x)$ where $\alpha_1 = \alpha^{-1}$. Applying again Proposition 2.6, but interchanging the ω_i's, we have an $\alpha_2 \in \mathcal{K}_\infty$ such that $\omega_2(x) \le \alpha_2(w_1(x))$ for all x, so (4) holds.

Conversely, suppose that (4) holds. Since $\omega_2(x) \geq \alpha_1(\omega_1(x))$ and $\alpha_1(\omega(x)) > 0$ for $x \notin \mathcal{A}$, it follows that $\omega_2(x) > 0$ for $x \notin \mathcal{A}$. On the other hand, $\omega_2(\mathcal{A}) \leq \alpha_2(w_1(\mathcal{A})) = 0$, so ω_2 is positive definite with respect to $\mathcal{A}$. It remains to show that ω_2 is proper. Pick any $r \geq 0$; we need to show that $S_r = \{x \mid \omega_2(x) \leq r\}$ is compact. Note that $S_r \subseteq S_r' := \{x \mid \omega_1(x) \leq \alpha_1^{-1}(r)\}$ and that the latter set is compact because ω_1 is proper. As the restriction of ω_2 to S_r' is continuous, S_r is closed in S_r' and therefore compact. ■

3. Systems with inputs

From now on, assume given an open subset $\mathbb{X} \subseteq \mathbb{R}^n$, $\mathcal{A} \subset \mathbb{X}$, and a size function ω for $(\mathbb{X}, \mathcal{A})$. We consider here systems with n state variables and m-dimensional inputs in the usual sense of control theory [5]:

$$\dot{x}(t) = f(x(t), u(t))$$

(the argument "t" is often omitted, and dot indicates derivative with respect to time). The map

$$f : \mathbb{X} \times \mathbb{R}^m \to \mathbb{R}^n$$

is assumed to be locally Lipschitz and

$$f(\mathcal{A}, 0) = 0.$$

States $x(t)$ take values in $\mathbb{X}$, and inputs (also called "controls" or "disturbances" depending on the context) are Lebesgue measurable essentially bounded maps

$$u : [0, \infty) \to \mathbb{R}^m.$$

We consider the sup norm of inputs:

$$\|u\|_\infty := \operatorname*{ess\,sup}_{t \geq 0} |u(t)|$$

where $|u|$ is the Euclidean norm in $\mathbb{R}^m$ and "ess sup" denotes essential supremum.

For each initial state x^0 and each input u, the solution of the initial value problem with initial state $x(0) = x^0$ and input u is denoted as

$$x(t, x^0, u) \in \mathbb{X}$$

and is defined on some maximal interval

$$[0, t_{\max}(x^0, u)).$$

Remark 3.1. For the sake of maximum generality, we allow inputs to be arbitrary (bounded) measurable functions. A technical issue is that measurable functions are in reality equivalence classes of functions, equal only up to measure zero subsets. Solutions of the differential equation are absolutely continuous functions and estimates over time have to be qualified by the phrase "for almost all t". We omit this qualification to make reading easier. In any event, for continuous inputs (which suffice for most applications) solutions are continuously differentiable and there is no need for the qualifier. □

3.1. Input to state stability

The notion of input to state stability (ISS), introduced in [3] (see expositions in [4–7]) provides a framework to describe stability features of the mapping $(x(0), u(\cdot)) \mapsto x(\cdot)$ that sends initial states and input functions into solution trajectories. Prominent among these features are that inputs that are bounded, small, "eventually" small, or convergent, should lead to states with the respective property. In addition, ISS quantifies how initial states affect transient behavior.

The formal definition, extended to open subsets, is as follows. Recall that a function $\beta : [0, \infty) \times [0, \infty) \to [0, \infty)$ is said to be of class $\mathcal{KL}$ if (1) for each fixed t, $\beta(s, t)$ as a function of r is in class $\mathcal{K}$ and (2) for each fixed r, $\beta(r, t)$ decreases to zero as $t \to \infty$.

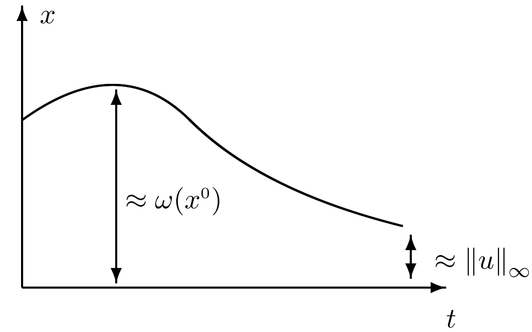

Fig. 1. ISS combines overshoot and asymptotic behavior.

Definition 3.2. A system is *input to state stable (ISS)* (on the open set $\mathbb{X}$ and with respect to $\mathcal{A}$) if there exist functions $\beta \in \mathcal{KL}$ and $\gamma \in \mathcal{K}_\infty$ so that the following property holds: for all inputs $u(\cdot)$ and all initial conditions $x^0 \in \mathbb{X}$, the solution is defined for all $t \geq 0$, that is, $t_{\max}(x^0, u) = +\infty$, and it satisfies the estimate:

$$\omega(x(t, x^0, u)) \leq \beta(\omega(x^0), t) + \gamma(\|u\|_\infty) \tag{ISS}$$

for all $t \geq 0$.

Note that this definition is independent of the particular size function used (although with β and γ functions that may change with ω) because of Proposition 2.6. When $\mathbb{X} = \mathbb{R}^n$, since $|x|_\mathcal{A}$ is a size function, this becomes the usual definition of ISS.

Since, in general, $\max\{a, b\} \leq a + b \leq \max\{2a, 2b\}$, one could restate the ISS condition in a slightly different manner, namely, asking for the existence of some $\beta \in \mathcal{KL}$ and $\gamma \in \mathcal{K}_\infty$ (in general different from the ones in the ISS definition) such that

$$\omega(x(t, x^0, u)) \leq \max\left\{\beta(\omega(x^0), t),\ \gamma(\|u\|_\infty)\right\}$$

holds for all solutions.

Intuitively, the definition of ISS requires that, for t large, the size of the state must be bounded by some function of the sup norm, that is to say, the maximum amplitude, of inputs, since $\beta(\omega(x^0), t) \to 0$ as $t \to \infty$. On the other hand, the term $\beta(\omega(x^0), 0)$ may dominate for small t, and this serves to quantify the magnitude of the transient (overshoot) behavior as a function of the size of the initial state x^0, see Fig. 1.

For stable (A having all eigenvalues with negative real part) linear systems $\dot{x} = Ax + Bu$ evolving on $\mathbb{X} = \mathbb{R}^n$, the variation of parameters formula gives immediately the following inequality:

$$|x(t)| \leq \beta(t)\left|x^0\right| + \gamma\|u\|_\infty,$$

where

$$\beta(t) = \|e^{tA}\| \to 0 \quad \text{and} \quad \gamma = \|B\| \int_0^\infty \|e^{sA}\| ds < \infty$$

(here $\|\cdot\|$ is induced operator norm). This is a particular case of the ISS estimate, $|x(t)| \leq \beta(|x^0|, t) + \gamma(\|u\|_\infty)$, with linear comparison functions. Note that $\beta(t) \leq Ce^{-\lambda t}$ for some $C > 0$ and some $\lambda > 0$, so one has exponential convergence when $u \equiv 0$.

Remark 3.3. We could think of a particular size function ω as an *output* function or "observable" $y = \omega(x)$ of the system $\dot{x} = f(x, u)$. With this interpretation, the definition is almost identical with that of "state-independent input to output stability" (SIIOS) given in [16]. The paper [16] presents a large number of results relating SIIOS to several other stability notions with respect to outputs. However, the interest in that paper is on *non-proper* ω, and $\mathbb{X}=\mathbb{R}^n$, since for proper functions and $\mathbb{X}=\mathbb{R}^n$, SIIOS would simply coincide with ISS. □

Remark 3.4. The definition of ISS on open sets is the only one that is invariant under coordinate changes and is also compatible with the definition of ISS on Euclidean spaces. To see this, let us take the case $\mathcal{A}=\{\bar{x}\}$. Generally, suppose that $\mathbb{X}\subseteq\mathbb{R}^n$ and $\mathbb{Y}\subseteq\mathbb{R}^n$ are open subsets (for example, $\mathbb{Y}=\mathbb{R}^n$), that $\varphi:\mathbb{X}\to\mathbb{Y}$ is a diffeomorphism with $\varphi(\bar{x})=\bar{y}$, and that $\omega_{\mathbb{Y}}$ is a size function for $(\mathbb{Y},\bar{y})$. Then $\omega_{\mathbb{X}}(x):=\omega_{\mathbb{Y}}(\varphi(x))$ defines a size function for $(\mathbb{X},\mathcal{A})$: continuity and positive definiteness are clear, and properness follows because, given any $r\geq 0$, the set $S_r:=\{x\mid\omega_{\mathbb{X}}(x)\leq r\}$ is the same as $\varphi^{-1}(\{y\mid\omega_{\mathbb{Y}}(y)\leq r\})$, and this set is compact because φ^{-1} is continuous and $\{y\mid\omega_{\mathbb{Y}}(Y)\leq r\}$ is compact. Now, given any system $\dot{x}=f(x,u)$ with state space $\mathbb{X}$, define the system $\dot{y}=g(y,u)$ with state space $\mathbb{Y}$ by mapping trajectories under φ, that is, $g(y,u):=\varphi_*(\varphi^{-1}(y))f(\varphi^{-1}(y),u)$, where φ_* denotes the Jacobian of φ. Suppose that this transformed system is ISS, so we have an estimate:

$$\omega_{\mathbb{Y}}(y(t,y^0,u))\ \leq\ \max\left\{\beta(\omega_{\mathbb{Y}}(y^0),t)\,,\ \gamma\left(\|u\|_\infty\right)\right\}$$

that holds for all solutions. For any initial state x^0 of the original system, $\varphi(x(t,x^0,u))=y(t,y^0,u)$, by construction. Since $\omega_{\mathbb{Y}}(\varphi(x(t,x^0,u)))=\omega_{\mathbb{X}}(x(t,x^0,u))$ by definition of $\omega_{\mathbb{X}}$, we conclude that the original system is ISS as well. Now, if a system is ISS on an open set with respect to a point equilibrium $\mathcal{A}$, there is a diffeomorphism $\varphi:\mathbb{X}\to\mathbb{Y}=\mathbb{R}^n$. Thus, the definition of ISS on Euclidean spaces maps precisely into the definition that we gave. □

3.2. ISS-Lyapunov functions

We now define ISS-Lyapunov functions on open sets. These are functions that can be used to certify that a system is ISS. We assume given a system $\dot{x}=f(x,u)$ as above.

Definition 3.5. A continuously differentiable $V:\mathbb{X}\to\mathbb{R}$ is said to be an *ISS-Lyapunov function* for $\dot{x}=f(x,u)$ with respect to $(\mathbb{X},\mathcal{A})$ if $V(x)=\overline{V}$ for $x\in\mathcal{A}$ and the following properties hold:

(a) $V-\overline{V}$ is a size function for $(\mathbb{X},\mathcal{A})$, and
(b) there exist functions $\alpha,\gamma\in\mathcal{K}_\infty$ such that

$$\dot{V}(x,u)\ \leq\ -\alpha(\omega(x))\ +\ \gamma(|u|)\quad\forall(x,u)\in\mathbb{X}\times\mathbb{R}^m \qquad \text{(L-ISS)}$$

where $\dot{V}:\mathbb{X}\times\mathbb{R}^m\to\mathbb{R}$ is the function:

$$\dot{V}(x,u):=\nabla V(x)\cdot f(x,u).$$

The interpretation of $\dot{V}$ is given by the fact that, for any solution $x(t)$ of $\dot{x}=f(x,u)$, the derivative $dV(x(t))/dt$ is $\dot{V}(x(t),u(t))$.

Remark 3.6. Property (a) in the definition of ISS-Lyapunov function is equivalent to the existence of two functions $\alpha_i\in\mathcal{K}_\infty$, $i=1,2$ such that

$$\alpha_1(\omega(x))\ \leq\ V(x)-\overline{V}\ \leq\ \alpha_2(\omega(x))\quad\forall x\in\mathbb{X}\,. \tag{5}$$

This is an immediate application of Corollary 2.7. Regarding property (b), redefining $\alpha:=\alpha\circ\alpha_2^{-1}\in\mathcal{K}_\infty$, one also has an estimate in which, instead of condition (L-ISS), one has the differential inequality:

$$\dot{V}(x,u)\ \leq\ -\alpha(V(x))-\overline{V}\ +\ \gamma(|u|)\quad\forall(x,u)\in\mathbb{X}\times\mathbb{R}^m\,. \qquad \text{(L-ISS')}$$

Conversely, suppose that (a) and (L-ISS') hold. Let α_1 be as in (5). Then $\widetilde{\alpha}(\omega(x))\leq\alpha(V(x))-\overline{V}$, where $\widetilde{\alpha}:=\alpha\circ\alpha_1$. This $\widetilde{\alpha}$ gives an estimate of the form (L-ISS). □

Theorem 1. *If a system admits an ISS-Lyapunov function, then it is ISS.* ■

This is a well-known fact, and its proof is entirely analogous to the proof for $\mathbb{X}=\mathbb{R}^n$ in the original paper [3]. We sketch the details here, starting from an estimate L-ISS'. Pick any solution $x(t,x^0,u)$, and define

$$v(t)\ :=\ V(x(t,x^0,u))-\overline{V}.$$

Note that $\dot{v}(t)=\dot{V}(x(t),u(t))\leq-\alpha(v(t))+\gamma(|u(t)|)$. For any t, either $\alpha(v(t))\leq 2\gamma(|u(t)|)$ or $\dot{v}(t)\leq-\alpha(v(t))/2$. From here, one deduces by a comparison theorem that

$$v(t)\ \leq\ \max\left\{\beta(v(0),t)\,,\ \alpha^{-1}(2\gamma(\|u\|_\infty))\right\}\quad\forall t\in[0,t_{\max}(x^0,u)),$$

where the $\mathcal{KL}$ function $\beta(s,t)$ is the solution $y(t)$ of the initial value problem

$$\dot{y}=-\frac{1}{2}\alpha(y)\,,\quad y(0)=s.$$

Using that $v(0)=V(x^0)-\overline{V}\leq\alpha_2(\omega(x^0))$ and $\omega(x(t,x^0,u))\leq\alpha_1^{-1}(V(x(t,x^0,u))-\overline{V})=\alpha_1^{-1}(v(t))$, we have

$$\begin{aligned}\omega(x(t,x^0,u))&\leq\max\left\{\alpha_1^{-1}(\beta(\alpha_2(\omega(x^0)),t))\,,\ \alpha_1^{-1}(\alpha^{-1}(2\gamma(\|u\|_\infty)))\right\}\\&\leq\max\left\{\widetilde{\beta}(\omega(x^0),t)\,,\ \widetilde{\gamma}(\|u\|)\right\}\ \forall t\in[0,t_{\max}(x^0,u))\,,\end{aligned}$$

with $\beta\in\mathcal{KL}$ and $\widetilde{\gamma}\in\mathcal{K}_\infty$. It only remains to prove that $t_{\max}(x^0,u)=+\infty$. To see this, note that, for any solution $x(t,x^0,u)$, we have the bound

$$\omega(x(t,x^0,u))\ \leq\ r:=\max\left\{\widetilde{\beta}(\omega(x^0),0)\,,\ \widetilde{\gamma}(\|u\|_\infty)\right\}.$$

Therefore, $x(t,x^0,u)\in S_r$ for all t on the maximal interval of definition of the solution. The set S_r is compact (properness of size functions), so the solution is defined for all $t\geq 0$ (see for example the ODE appendix in [5]).

The converse of Theorem 1 is also true: *if a system is ISS, then it admits an ISS-Lyapunov function.* Again, the proof is entirely analogous to that for the case $\mathbb{X}=\mathbb{R}^n$, proved in [17,18], which is basically a theorem about Lyapunov functions for differential inclusions. We have not been able to find a clean citation for this converse result, though it can be easily derived from a far more general theorem, valid for hybrid systems, given in [13]. In the special case of a singleton $\mathcal{A}=\{\bar{x}\}$, one can derive the theorem from the case $\mathbb{X}=\mathbb{R}^n$ as follows. If a system is ISS, then the system with zero inputs $\dot{x}=f(x,0)$ has $\bar{x}$ as an asymptotically stable point with domain of attraction all of $\mathbb{X}$ (this follows from the estimate $\omega(x)\leq\beta(\omega(0),t)$). This implies that $\mathbb{X}$ is diffeomorphic to $\mathbb{R}^n$, see Theorem 2.2 in [19], which obtains this as a simple corollary of the Brown–Stallings Theorem. (A proof of a simpler fact, that $\mathbb{X}$ must be contractible, is very easy; see for example theorem 21 in [5].) This means that under a diffeomorphism, we can apply the result for $\mathbb{X}=\mathbb{R}^n$, and when transforming back, we obtain an L-ISS Lyapunov function.

4. Application to gradient systems

We assume given a pair $(\mathbb{X},\mathcal{A})$ and a size function ω for $(\mathbb{X},\mathcal{A})$. We write the gradient of a function $V:\mathbb{X}\to\mathbb{R}$ as a row (co)vector ∇V, and its Euclidean norm as $|\nabla V|$. When ∇V is locally Lipschitz, the gradient flow has unique solutions and if ∇V is globally Lipschitz, these solutions are automatically defined for all $t\geq 0$. (In the notations of Nesterov's book [20], the set of functions V for which ∇V has a uniform Lipschitz constant L is denoted $C_L^{1,1}(\mathbb{X})$. In our setup, solutions are defined for all $t\geq 0$ even if ∇V is not assumed to be globally Lipschitz.)

4.1. Proper loss functions

Definition 4.1. A continuously differentiable $V : \mathbb{X} \to \mathbb{R}$, with locally Lipschitz continuous gradient ∇V, will be said to be a *proper loss function* with respect to $(\mathbb{X}, \mathcal{A})$ if $V(x) = \overline{V}$ for $x \in \mathcal{A}$ and the following properties hold:

(a) $V - \overline{V}$ is a size function for $(\mathbb{X}, \mathcal{A})$, and
(b) $|\nabla V|$ is a size function for $(\mathbb{X}, \mathcal{A})$. □

Lemma 4.2. *Suppose given a continuously differentiable $V : \mathbb{X} \to \mathbb{R}$, with locally Lipschitz continuous gradient ∇V, so that $V(x) = \overline{V}$ for $x \in \mathcal{A}$ and such that $V - \overline{V}$ is a size function for $(\mathbb{X}, \mathcal{A})$. Then these two properties are equivalent:*

- *V is a proper loss function,*
- *for some $\alpha \in \mathcal{K}_\infty$,*

$$\alpha(\omega(x)) \;\leq\; |\nabla V(x)| \qquad \text{for all } x \in \mathbb{X}.$$

Proof. Suppose that V is a proper loss function. Since $|\nabla V|$ is a size function for $(\mathbb{X}, \mathcal{A})$, there is an $\alpha \in \mathcal{K}_\infty$ as claimed, by Corollary 2.7 (take $\alpha = \alpha_1$). Conversely, suppose that $\alpha(\omega(x)) \leq |\nabla V(x)|$ with $\alpha \in \mathcal{K}_\infty$. Then $\nabla V(\mathcal{A}) = 0$, because V has a (local and even global) minimum at $\mathcal{A}$. For $x \notin \mathcal{A}$, $0 < \alpha(\omega(x)) \leq |\nabla V(x)|$, so $|\nabla V(x)|$ is positive definite.

It remains to show that $|\nabla V|$ is proper. Pick any $r \geq 0$; we need to show that $S_r = \{x \mid |\nabla V(x)| \leq r\}$ is compact. Note that $S_r \subseteq S'_r := \{x \mid \omega(x) \leq \alpha^{-1}(r)\}$ and that the latter set is compact because ω is proper. As the restriction of ∇V to S'_r is continuous, S_r is closed in S'_r and therefore compact. ■

Applied with $\omega = V - \overline{V}$, Lemma 4.2 together with the definition of size function says that an equivalent way to define a proper loss function is to ask:

- V is continuously differentiable, with locally Lipschitz continuous gradient;
- V achieves a strict global minimum at $x \in \mathcal{A}$;
- V is proper; and
- there is some $\alpha \in \mathcal{K}_\infty$ such that

$$\alpha(V(x) - \overline{V}) \;\leq\; |\nabla V(x)|^2 \qquad \text{for all } x \in \mathbb{X}\,. \tag{6}$$

We wrote $|\nabla V(x)|^2$ instead of $|\nabla V(x)|$ for convenience in what follows; this makes no difference, since $(\alpha(\cdot))^2$ is a $\mathcal{K}_\infty$ function if and only if α is.

In many problems one can directly obtain an estimate as in (6), and this is useful for obtaining explicit ISS stability rates.

4.2. Gradient flow is ISS

Fix a proper loss function V, a constant $\eta > 0$ (the "learning rate"), and a locally Lipschitz mapping $B : \mathbb{X} \to \mathbb{R}^{n\times m}$ with bounded image. We consider the gradient system in (1), repeated here for convenience:

$$\dot{x}(t) \;=\; -\eta\, \nabla V(x(t))^T \;+\; B(x(t))u(t).$$

Theorem 2. *If V is a proper loss function, then system (1) is ISS.*

Proof. We will prove that $V - \overline{V}$ is an ISS-Lyapunov function for (1). Since $V - \overline{V}$ is a size function, we need to show an estimate (L-ISS'). We have:

$$\begin{aligned}
\dot{V}(x, u) &= -\eta|\nabla V(x)|^2 \;+\; \nabla V(x)\, B(x)u \\
&= -\eta|\nabla V(x)|^2 + (\sqrt{\eta}\nabla V(x))(\sqrt{1/\eta}B(x)u) \\
&\leq -\eta|\nabla V(x)|^2 + \frac{\eta}{2}|\nabla V(x)|^2 + \frac{1}{2\eta}|B(x)u|^2 \\
&= -\frac{\eta}{2}|\nabla V(x)|^2 + \frac{1}{2\eta}|B(x)u|^2 \\
&\leq -\frac{\eta}{2}\alpha(V(x) - \overline{V}) + \frac{K_B}{2\eta}|u|^2 \\
&= -\widetilde{\alpha}(V(x) - \overline{V}) + \gamma(|u|)
\end{aligned}$$

where we used that, for row and column vectors in $\mathbb{R}^n$, $|vw| \leq |v||w| \leq (1/2)(|v|^2 + |w|^2)$ (Cauchy–Schwarz inequality followed by $2ab \leq a^2 + b^2$), and the inequality $\alpha(V(x) - \overline{V}) \leq |\nabla V(x)|^2$, and where K_B is an upper bound on the induced Euclidean norms $|B(x)|$, $x \in \mathbb{X}$, and defined $\widetilde{\alpha} := \frac{\eta}{2}\alpha \in \mathcal{K}_\infty$ and $\gamma := \frac{K_B}{2\eta}r^2 \in \mathcal{K}_\infty$. So V is an ISS-Lyapunov function, and thus the system (1) is ISS. ■

In the particular case in which the estimate $\alpha(V(x) - \overline{V}) \leq |\nabla V(x)|^2$ holds with a linear function α, the proof of Theorem 1 provides a rate of decrease for $v(t) = V(x(t)) - \overline{V}$ which is exponential: the function $\beta(r, t)$ has the form $e^{-\lambda t}r$ for some positive λ.

4.3. An example: LQR problem

The (infinite-horizon) LQR problem is one of the best-studied optimal control problems. Consider a time-invariant linear system

$$\dot{x} \;=\; Ax + Bu$$

and define the cost function:

$$\mathcal{J}(x^0, u) \;:=\; \int_0^\infty x^T(t)Qx(t) \;+\; u^T(t)Ru(t)\, dt$$

where $x(t) = x(t, x^0, u)$. Here $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ are matrices so that the pair (A, B) is controllable (or even just stabilizable or "asymptotically controllable"), which guarantees the finiteness of the objective function, and $Q \in \mathbb{R}^{n\times n}$, and $R \in \mathbb{R}^{m\times m}$ are positive definite. The objective is to minimize $\mathcal{J}(x^0, u)$ over all measurable essentially bounded control functions $u : [0, \infty) \to \mathbb{R}^m$, for any given x^0.

The unique optimal control is obtained by using the linear feedback law $u(t) = -Kx(t)$, where $K = R^{-1}B^T\Pi$ and Π is the unique positive definite solution of the algebraic Riccati equation

$$\Pi BR^{-1}B^T\Pi \;-\; A^T\Pi \;-\; \Pi A \;-\; Q \;=\; 0$$

(that is, $u(t) = -Kx(t)$), where x solves $\dot{x} = (A - BK)x$ with $x(0) = x^0$), and at this optimum value,

$$\mathcal{J}(x^0, u) \;=\; (x^0)^T\Pi x^0$$

(see, for instance, Theorem 41 in [5]). The optimal feedback matrix $K = R^{-1}B^T\Pi$ stabilizes the system, i.e., $A - BK$ is a Hurwitz matrix (all eigenvalues have negative part).

Since the optimal control is given by a linear feedback, one may pose the simpler question of optimizing over all feedback matrices which belong to the open set $\mathbb{X} := \{K \mid A - BK$ is Hurwitz$\}$. In terms of K and using $u = -Kx$, one can introduce the loss function

$$V_{x^0}(K) \;:=\; \int_0^\infty x(t)^TQx(t) \;+\; (-Kx(t))^T\, R\,(-Kx(t))\, dt$$

where $x(t)$ solves $\dot{x} = (A - BK)x$, i.e., $x(t) = e^{(A-BK)t}x^0$, so that we can also write

$$\begin{aligned}
V_{x^0}(K) &= \int_0^\infty x(t)^T(Q + K^TRK)x(t)\, dt \\
&= \operatorname{trace}\left((Q + K^TRK)\int_0^\infty x(t)x(t)^T\, dt\right)
\end{aligned}$$

where we have used that for a scalar $a = \mathrm{trace}(a)$ and that $\mathrm{trace}(UV) = \mathrm{trace}(VU)$. To obtain a simpler problem, we assume now that the initial state is picked distributed randomly according to some probability density in $\mathbb{R}^n$ (for example, Gaussian) with covariance $\Sigma = \mathbb{E}[x^0(x^0)^T]$ and we wish to minimize

$$V(K) := \mathbb{E}[V_{x^0}] = \mathrm{trace}((Q + K^T RK)P)$$

with

$$P = \mathbb{E}\left[\int_0^\infty e^{tF} x^0(x^0)^T e^{tF^T}\, dt\right] = \int_0^\infty e^{tF}\, \Sigma\, e^{tF^T}\, dt$$

where $F = A - BK$. It follows (see for instance Theorem 18 in [5]) that P is the (unique) solution of the Lyapunov matrix equation

$$(A - BK)P + P(A - BK)^T + \Sigma = 0 \tag{7}$$

In summary, one has to minimize the loss function $V(K) = \mathrm{trace}((Q + K^T RK)P)$ where the positive definite matrix P satisfies (7) and $K \in \mathbb{X}$. Since the solution P of the linear system of Eqs. (7) is a rational function of the data (Cramer's rule), it follows that $V(K)$ is rational in the entries of the matrix K, and hence V is differentiable. Although it is not generally convex, it has a unique global minimum at the optimal $K = R^{-1}B^T\Pi$. It is also known that it is a proper function, see [21]. The gradient can be computed as follows (this is implicit in the computations in [21,22], but see [23] for a clear exposition):

$$\nabla V(K) = 2(RK - B^T L)P$$

where L is the unique positive definite matrix that satisfies

$$(A - BK)^T L + L(A - BK) + Q + K^T RK = 0.$$

For example, suppose that $n = m = 1$, $a = q = r = \Sigma = 1$. In this case $\mathbb{X} = \{k \mid bk > 1\}$ and one obtains

$$V(k) = \frac{k^2 + 1}{2(bk - 1)}$$

and

$$V'(k) = \frac{bk^2 - 2k - b}{2(bk - 1)^2}.$$

In general, it can be shown, see [1], that V is a proper loss function. In fact, that reference shows that the Polyak–Łojasiewicz condition [24]

$$c_r(V(x) - \overline{V}) \le |\nabla V(x)|^2$$

holds on sublevel sets, for constants c_r, which implies that a lower bounding $\alpha \in \mathcal{K}_\infty$ exists.

4.4. The condition that $|\nabla V|$ be a size function is key

Weaker positive definiteness conditions than (6) may still result in global convergence of gradient flows but may not guarantee that the gradient system with inputs (1) is ISS. We illustrate this with an example.

Take $\mathbb{X} = (0, \infty) \subset \mathbb{R}$, $V(x) = x + \frac{1}{x}$, $\mathcal{A} = \{1\}$, $\overline{V} = 2$. Then $V - \overline{V}$ is a size function with respect to $(\mathbb{X}, \mathcal{A})$. The associated gradient system with inputs is

$$\dot{x} = \frac{1}{x^2} - 1 + u$$

which, when the input is $u \equiv 0$ has 1 as a globally asymptotically stable system. In other words, with no disturbances in the gradient calculation, there is global convergence to the global minimum, as expected since V is convex. However, when $u \equiv 1$ we obtain the equation $\dot{x} = \frac{1}{x^2}$, whose solutions diverge to $+\infty$.

Observe that $|\nabla V|$ is not a size function, since $|1 - 1/x^2| \to 1$ as $x \to +8$.

5. ISS and steepest descent

In this section, we fix a pair $(\mathbb{X}, \mathcal{A})$ and a size function ω for $(\mathbb{X}, \mathcal{A})$.

We recall from the introductory discussion that we are interested in proving that the steepest descent iteration $x^+ = x - \lambda[\nabla V(x)^T + B(x)u]$, where λ is picked at each step of the iteration so as to minimize the value $V(x^+)$, is a discrete-time ISS system. We start by reviewing some simple properties of Lipschitz functions.

5.1. Gradients of locally lipschitz functions on $\mathbb{X}$

Suppose given a continuously differentiable function $V : \mathbb{X} \to \mathbb{R}$ such that these two properties hold:

[SV] $V - \overline{V}$ is a size function for $(\mathbb{X}, \mathcal{A})$;
[LL] ∇V is locally Lipschitz.

We next review a couple of well-known facts about Lipschitz functions.

Remark 5.1. For each compact subset $K \subset \mathbb{X}$, there some $L \ge 0$ such that the one-sided Lipschitz estimate

$$(\nabla V(y) - \nabla V(x))\,(y - x) \le L\,|y - x|^2 \tag{8}$$

holds for all $x, y \in K$. Indeed, the function ∇V is Lipschitz on K, with some constant L (start locally and take finite subcovers), so

$$|(\nabla V(y) - \nabla V(x))\,(y - x)| \le |\nabla V(y) - \nabla V(x)|\,|y - x|$$
$$\le (L\,|y - x|)\,|y - x| = L\,|y - x|^2$$

by the Cauchy–Schwarz inequality and the Lipschitz property. □

Remark 5.2. Suppose that $x, y \in K$ and that L is a one-sided Lipschitz constant as in (8) on the segment

$$K = [x, y] := \{z \mid z = x + s(y - x), s \in [0, 1]\}$$

connecting x and y. Then

$$V(y) \le V(x) + \nabla V(x)(y - x) + \frac{L}{2}\,|y - x|^2\,. \tag{9}$$

This is a standard fact, see e.g. [20]. The blanket assumption $\mathbb{X} = \mathbb{R}^n$ made there is not needed; since the proof is so simple, we write it here. Pick x, y, and consider the continuously differentiable function:

$$g : [0, 1] \to \mathbb{R} : s \mapsto V(x + s(y - x)).$$

Then

$$\begin{aligned} V(y) - V(x) - \nabla V(x)(y - x) &= g(1) - g(0) - g'(0) \\ &= \int_0^1 g'(s)\,ds - g'(0) \\ &= \int_0^1 [g'(s) - g'(0)]\,ds \end{aligned}$$

where

$$\begin{aligned} g'(s) - g'(0) &= \nabla V(x + s(y - x))(y - x) - \nabla V(x)(y - x) \\ &= \frac{1}{s}\,[\nabla V(x + s(y - x)) - \nabla V(x)]\,(s(y - x)) \\ &\le sL\,|y - x|^2 \end{aligned}$$

by (8) when $s \ne 0$ (and this is trivial when $s = 0$). Therefore

$$V(y) - V(x) - \nabla V(x)(y - x) \le \int_0^1 sL\,|y - x|^2\, ds = \frac{L}{2}\,|y - x|^2\,,$$

as desired. □

We wish to study the behavior of steepest descent when the gradient of V is inaccurately estimated.

From now on we assume that ∇V is positive definite:

[PD] $\nabla V(x) \neq 0$ for all $x \notin \mathcal{A}$.

in addition to **[SV]** and **[LL]**.

Lemma 5.3. *Pick any $x^0 \in \mathbb{X}$, $x^0 \notin \mathcal{A}$, and let L be a Lipschitz constant for ∇V on the compact set*

$$S := \{x \in \mathbb{X} \mid V(x) \le V(x^0)\}$$

(without loss of generality, $L > 0$). Pick any $q \in \mathbb{R}^n$ and write $p := \nabla V(x^0)^T \neq 0$. Suppose that $\lambda > 0$ has the property that

$$x^0 - \mu(p+q) \in S \quad \text{for each } 0 \le \mu \le \lambda.$$

Then

$$V(x^0 - \lambda(p+q)) - V(x^0) \le \left(-\lambda + \frac{\lambda^2 L}{2}\right)|p|^2 + \frac{\lambda^2 L}{2}|q|^2 + \left(\lambda + \lambda^2 L\right)|p|\,|q| \,. \tag{10}$$

Proof. Let $x = x^0$ and $y = x^0 - \lambda(p+q)$. The segment $[x, y]$ consists of points of the form $x^0 - \mu(p+q)$, with $0 \le \mu \le \lambda$. Therefore, we may apply the Lipschitz estimate (9), to obtain:

$$V(x^0 - \lambda(p+q)) - V(x^0) \;\le\; -\lambda p^T(p+q) + \frac{\lambda^2 L}{2}|p+q|^2 \,.$$

Since

$$|p+q|^2 \;=\; |p|^2 + |q|^2 + 2p^T q \;\le\; |p|^2 + |q|^2 + 2\,|p|\,|q|$$

and similarly $-p^T q \;\le\; \left|p^T q\right| \;\le\; |p|\,|q|$, the estimate (10) follows. ■

We have this immediate consequence:

Corollary 5.4. *Suppose that $q \le c\,|p|$ in Lemma 5.3. Then,*

$$V(x^0 - \lambda(p+q)) - V(x^0) \;\le\; \lambda\left[(c-1) + \frac{\lambda L}{2}(c+1)^2\right]|p|^2 \,.$$

In particular, taking $c = 1/2$ and $\lambda \le \frac{2}{9L}$, then $V(x^0 - \lambda(p+q)) - V(x^0) \le -\frac{\lambda}{4}|p|^2$. □

Lemma 5.5. *Pick x^0, L, q, and p as in Lemma 5.3. Suppose that $|q| \le \frac{1}{2}|p|$ and $\lambda = \frac{2}{9L}$. Then $x^0 - \mu(p+q) \in \mathbb{X}$ for each $0 \le \mu \le \lambda$ and*

$$V(x^0 - \lambda(p+q)) - V(x^0) \;\le\; -\frac{1}{18L}|p|^2 \,.$$

Proof. Since $\mathbb{X}$ is an open set, $x^0 - \mu(p+q) \in \mathbb{X}$ for all small $\mu > 0$. Also, since

$$\left.\frac{d}{ds}\right|_{s=0} V(x^0 - s(p+q)) = -p^T(p+q) \;=\; -|p|^2 + p^T q \le -|p|^2 + |p|\,|q| \le -\frac{1}{2}|p|^2 \;<\; 0,$$

there is some $\varepsilon > 0$ such that

$$V(x^0 - \mu(p+q)) \;<\; V(x^0) \quad \text{for all } \mu \in (0, \varepsilon).$$

Suppose that there would exist some $\mu \in [0, \lambda]$ such that $x^0 - \mu(p+q) \notin S$. Since S is compact and $\mathbb{X}$ is open, this would mean that there is some $\mu \in [\varepsilon, \lambda]$ such that $x^0 - \mu(p+q) \in \mathbb{X}$ and $V(x^0 - \mu(p+q)) = V(x^0)$. To apply Corollary 5.4, we need to see that this cannot happen. Let

$$\lambda_0 \;:=\; \min\{\mu \in [\varepsilon, \lambda] \mid V(x^0 - \mu(p+q)) = V(x^0)\} \;\ge\; \varepsilon \;>\; 0.$$

Since $V(x^0 - \mu(p+q)) \le V(x^0)$ for all $\mu \in [0, \lambda_0]$, we may apply Corollary 5.4 to λ_0 to conclude that

$$0 \;=\; V(x^0 - \lambda(p+q)) - V(x^0) \;\le\; -\frac{\lambda_0}{4}|p|^2 \,,$$

which contradicts $\lambda_0 > 0$ and $p \neq 0$. Thus the hypotheses of Lemma 5.3 hold, and applying Corollary 5.4 to λ we conclude that

$$V(x^0 - \lambda(p+q)) - V(x^0) \;\le\; -\frac{\lambda}{4}|p|^2 \;=\; -\frac{1}{18L}|p|^2$$

as claimed. ■

5.2. Line search in direction of steepest descent

We continue with the assumptions **[PD], [SV], [LL]** on V. We next define a function

$$F : \mathbb{X} \times \mathbb{R}^n \to \mathbb{X}$$

that will represent an individual steepest descent step when starting at a point $x^0 \in \mathbb{X}$ and the (transpose of the) gradient is estimated as $p + q$ where $p := \nabla V(x^0)^T$ and $q \in \mathbb{R}^n$ represents an additive noise. Now take any $x^0 \in \mathbb{X}$ and any $q \in \mathbb{R}^n$ such that $p + q \neq 0$. Define

$$\Lambda(x^0, q) \;:=\; \left\{\lambda \ge 0 \mid V(x^0 - \mu(p+q)) \le V(x^0) \text{ for all } \mu \in [0, \lambda]\right\}.$$

Note that $0 \in \Lambda(x^0, q)$.

Suppose that $x^0 \notin \mathcal{A}$. We claim that the set $\Lambda(x^0, q)$ is compact. It is bounded above: otherwise, it would be the case that $x^0 - \lambda(p+q) \in \mathbb{X}$ and $V(x^0 - \lambda(p+q)) \le V(x^0)$ for all $\lambda \ge 0$; then since V is proper, the set of points $x^0 - \lambda(p+q)$ is bounded, but this contradicts that $\left|x^0 - \lambda(p+q)\right| \ge \left|\left|x^0\right| - \lambda\,|p+q|\right| \to \infty$ as $\lambda \to \infty$ because $|p+q| \neq 0$. It is also closed. Indeed, suppose that $\lambda_k \to \lambda$, with $\lambda_k \in \Lambda(x^0, q)$. Then $V(x^0 - \lambda(p+q)) \le V(x^0)$, by continuity. In addition, for each $\mu < \lambda$, there is some k so that $\mu < \lambda_k$ so $V(x^0 - \mu(p+q)) \le V(x^0)$, proving that $\lambda \in \Lambda(x^0, q)$. Thus we may define

$$\bar{\lambda}_{x^0,q} \;:=\; \operatorname*{argmin}_{\lambda \in \Lambda(x^0,q)} V(x^0 - \lambda(p+q))$$

where "arg min" means that we take the smallest λ that achieves this minimum value in the direction of $p + q$ when there is more than one. We then define

$$F(x^0, q) \;:=\; x^0 - \bar{\lambda}_{x^0,q}(p+q).$$

and $F(x^0, q) := x^0$ if $p + q = 0$ or if $x^0 \in \mathcal{A}$.

Note that $V(F(x^0, q)) \le V(x^0)$, because $0 \in \Lambda(x^0, q)$ and we are minimizing. In other words,

$$\widetilde{\Delta}V(x^0, q) \;:=\; V(F(x^0, q)) - V(x^0) \;\le\; 0 \quad \forall (x^0, q)$$

and observe that $\widetilde{\Delta}V(x^0, q) = 0$ if $V(x^0 - \varepsilon(p+q)) \ge V(x^0)$ for all small ε. On the other hand, since

$$\left.\frac{d}{ds}\right|_{s=0} V(x^0 - sp) \;=\; -\left|\nabla V(x^0)\right|^2 < 0$$

it follows that $\widetilde{\Delta}V(x^0, 0) < 0$ for all $x^0 \notin \mathcal{A}$.

We next estimate $\widetilde{\Delta}V(x^0, q)$ for all $|q|$ that are not "too large" compared to $|p|$.

Suppose that L is any Lipschitz constant for ∇V on the set $S = \{x \in \mathbb{X} \mid V(x) \le V(x^0)\}$, $|q| \le \frac{1}{2}|p|$, and $\lambda = \frac{2}{9L}$. From Lemma 5.5, $\lambda \in \Lambda(x^0, q)$, so $V(F(x^0, q)) \le V(x^0 - \lambda(p+q))$ by definition of $\bar{\lambda}_{x^0,q}$ as a minimizer. Thus, again by the Lemma,

$$\widetilde{\Delta}V(x^0, q) = V(F(x^0, q)) - V(x^0) \;\le\; V(x^0 - \lambda(p+q)) - V(x^0) \le -\frac{1}{18L}\left|\nabla V(x^0)\right|^2 \,. \tag{11}$$

5.3. Steepest descent with inputs

We now consider a slightly more general setup as follows. Let $B : \mathbb{X} \to \mathbb{R}^{n\times m}$ be a mapping with a bounded image, let K_B be an upper bound on the induced Euclidean norms $\{\|B(x)\|,\ x \in \mathbb{X}\}$. Assume that the gradient error at each iteration step is $q = B(x)u$, where $u \in \mathbb{R}^m$; thus $|q| \le K_B\,|u|$.

We define the steepest descent algorithm, with inputs u, as the discrete-time system defined by the following iteration function $f : \mathbb{X} \times \mathbb{R}^m \to \mathbb{X}$:

$$x^+ \;=\; f(x,u) := F(x, B(x)u).$$

We define $\Delta V(x,u) := \widetilde{\Delta}V(x, B(x)u)$, that is

$$\Delta V(x,u) \;:=\; V(f(x,u)) - V(x).$$

Since $\widetilde{\Delta}V(x,q) \le 0$ for all (x,q), also $\Delta V(x,u) \le 0$ for all (x,u). Obviously, we can also write

$$\Delta V(x,u) \;=\; [V(f(x,u)) - \overline{V}] - [V(x) - \overline{V}].$$

which exhibits ΔV as the change, in each steepest descent step, of the "excess cost" of V compared to its minimum value $\overline{V}$.

5.4. Discrete-time ISS

We now review input to state stability for discrete time systems (on open subsets), a notion which is completely analogous to that for continuous time, see for instance [25–27]. We consider discrete-time systems $x^+ = f(x,u)$, where $f : \mathbb{X} \times \mathbb{R}^m \to \mathbb{X}$ is a continuous function and $f(x,0) = x$ for $x \in \mathcal{A}$. Again, ω is a size function.

Definition 5.6. The discrete-time system $x^+ = f(x,u)$ is *input to state stable (ISS)* (on the open set $\mathbb{X}$ and with respect to $\mathcal{A}$) if there exist functions $\beta \in \mathcal{KL}$ and $\gamma \in \mathcal{K}_\infty$ so that the following property holds: for all input sequences $u = (u_0, u_1, \ldots) \in \ell^\infty_m$ and all initial conditions $x^0 \in \mathbb{X}$, the solution x^0 satisfies the estimate:

$$\omega(x(t,x^0,u)) \;\le\; \beta(\omega(x^0),t) \;+\; \gamma\left(\|u\|_\infty\right) \tag{ISS}$$

for all $t = 0, 1, 2, \ldots$.

Here $\|u\|_\infty = \sum_{t=0}^{\infty} |u_t|$ and $x(t,x^0,u)$ is obtained by solving recursively $x_{t+1} = f(x_t, u_t)$.

There are several equivalent definitions of ISS-Lyapunov function for discrete time systems. We pick here the most convenient one for the current application.

For any function $V : \mathbb{X} \to \mathbb{R}$, we denote $\Delta V(x,u) := V(f(x,u)) - V(x)$.

Definition 5.7. A continuous $V : \mathbb{X} \to \mathbb{R}$ is said to be an *ISS-Lyapunov function* for $x^+ = f(x,u)$ with respect to $(\mathbb{X}, \mathcal{A})$ if $V(x) = \overline{V}$ for $x \in \mathcal{A}$ and the following properties hold:

(a) $V - \overline{V}$ is a size function for $(\mathbb{X}, \mathcal{A})$, and
(b) there exist (i) a continuous and positive definite function α, and (ii) a function $\chi \in \mathcal{K}_\infty$, such that:

$$\omega(x) \ge \chi(|u|) \;\Rightarrow\; \Delta V(x,u) \le -\alpha(V(x) - \overline{V}) \tag{12}$$

for all $x \in \mathbb{X}$, $u \in \mathbb{R}^m$.

Equivalences among alternative ISS-Lyapunov function definitions, including a condition of the type $\Delta V(x,u) \le -\alpha(\omega(x)) + \gamma(|u|)$ for functions of class $\mathcal{K}_\infty$, are discussed in Remark 3.3 of [26]. As with continuous-time systems, the existence of ISS-Lyapunov functions is equivalent to the ISS property, see [13, 25,27]. For completeness, and because of the interest in open subsets $\mathbb{X}$, we prove the sufficiency below, appealing to some key technical lemmas in [26,27].

Let us write, for simplicity of notation, $W(x) = V(x) - \overline{V}$. As $\Delta V(x,u) = [V(f(x,u)) - \overline{V}] - [V(x) - \overline{V}] = W(f(x,u)) - W(x)$, one can write (12) as:

$$\omega(x) \ge \chi(|u|) \;\Rightarrow\; W(f(x,u)) \le W(x) - \alpha(W(x)).$$

Theorem 3. *If a discrete-time system admits an ISS-Lyapunov function V then it is ISS.*

Proof. We first remark that one may redefine V, replacing it by a function of the form $\rho(V(x))$ with $\rho \in \mathcal{K}_\infty$, in such a manner that the estimate (12) holds but now $\alpha \in \mathcal{K}_\infty$ (and the redefined V is so that $V - \overline{V}$ still a size function). The argument is similar to the one given in [3] for the continuous time case, but it is more delicate, see the proof of Lemma 2.8 in [27]. Moreover, one may assume that $r \mapsto r - \alpha(r)$ is of class $\mathcal{K}$ (see Lemma B.1 in [26]). So, from now on, and redefining V in this manner if needed, we will assume that α satisfies these two properties. Since W is a size function, there is a $\pi \in \mathcal{K}_\infty$ such that $\omega(x) \ge \pi(W(x))$, and thus $W(x) \ge \pi^{-1}(\chi(|u|))$ implies $\omega(x) \ge \chi(|u|)$, so redefining χ as $\pi^{-1} \circ \chi$ we can state the ISS-Lyapunov property as:

$$W(x) \ge \chi(|u|) \;\Rightarrow\; W(f(x,u)) \le W(x) - \alpha(W(x)).$$

Now let $\beta(r,t)$ be the solution of the scalar difference equation

$$y_{t+1} \;=\; y_t - \alpha(y_t), \quad y_0 = r \ge 0.$$

The property that $r \mapsto r - \alpha(r)$ is of class $\mathcal{K}$ implies $y_t \ge 0$ for all t, and also that $y_t < y'_t$ implies $y_{t+1} < y'_{t+1}$ for any two solutions, in other words, the iteration is monotone (it preserves order). Thus the function β is of class $\mathcal{K}$ on r. Moreover, since $\alpha(y) \ge 0$, $y_{t+1} \le y_t$, iterates form a decreasing sequence. Thus all solutions converge to zero as $t \to \infty$, since the only equilibrium $y - \alpha(y) = y$ is at $y = 0$. So $\beta \in \mathcal{KL}$.

We introduce the following function $\gamma : [0,\infty) \to [0,\infty)$:

$$\gamma(\mu) \;:=\; \max\{W(f(x,u)) \mid |u| \le \mu,\, W(x) \le \chi(\mu)\}$$

which is well-defined (the set over which we are maximizing is compact, and $W(f(x,u))$ is continuous on (x,u)), nondecreasing (the sets are larger as μ increases), and satisfies $\gamma(0) = 0$ (since $W(x) = 0$ implies $x \in \mathcal{A}$ and $f(x,0) = x$ for $x \in \mathcal{A}$). Note that this implication holds:

$$W(x) \le \chi(|u|) \;\Rightarrow\; W(f(x,u)) \le \gamma(|u|).$$

Replacing γ by a larger function if needed, we may assume that $\gamma \in \mathcal{K}_\infty$ and also that $\gamma(\mu) \ge \chi(\mu)$ for all μ. Consider the following sets:

$$P_\mu \;:=\; \{x \mid W(x) \le \gamma(\mu)\}\,.$$

We claim that this set is forward invariant for inputs with $\|u\|_\infty \le \mu$. Indeed, pick any $x \in P_\mu$ and any $u \in \mathbb{R}^m$ with $\|u\|_\infty \le \mu$. If $W(x) \ge \chi(|u|)$, then $W(f(x,u)) \le W(x) \le \gamma(\mu)$, so $f(x,u) \in P_\mu$. If instead $W(x) \le \chi(|u|)$, then $W(f(x,u)) \le \gamma(|u|) \le \gamma(\mu)$ as well.

Consider now any input u, any initial state x^0, and the corresponding solution $x(t,x^0,u)$ of $x^+ = f(x,u)$. Let $a_t := W(x(t,x^0,u))$ for $t = 0, 1, \ldots$, and $\mu := \|u\|_\infty$. We will compare this sequence to $y_t = \beta(W(x^0),t) = \beta(a_0,t)$. Note that by definition $a_0 = y_0$.

Consider first the case that $a_t \le \gamma(\mu)$ for all t. Obviously in that case $a_t \le \max\{\beta(a_0,t), \gamma(\mu)\}$ for all t.

Consider next the case that $a_t > \gamma(\mu)$ for some t. Then either (i) $a_t > \gamma(\mu)$ for all t, or (ii) there is a $T \ge 0$ so that $a_t \ge \gamma(\mu)$ for $t = 0, \ldots, T$ and $a_{T+1} \le \gamma(\mu)$. Suppose that $a_t \ge \gamma(\mu)$ for

$t = 0, \ldots, T$. From the ISS-Lyapunov property, we know that

$$a_t \geq \gamma(\mu) \quad \Rightarrow \quad a_t \geq \chi(\mu) \geq \chi(|u_t|)$$
$$\Rightarrow \quad a_{t+1} \leq a_t - \alpha(a_t).$$

We claim that $a_t \leq y_t = \beta(a_0, t)$ for $t = 0, \ldots, T$. This holds for $t = 0$. In general, if $\gamma(\mu) \leq a_t \leq y_t$ then $a_{t+1} \leq a_t - \alpha(a_t) \leq y_t - \alpha(y_t) = y_{t+1}$, because $r - \alpha(r)$ is nondecreasing in r. By induction, $a_t \leq y_t$ for $t = 0, \ldots, T$. It cannot be that (i) holds, since $y_t \to 0$ as $t \to \infty$. Thus (ii) holds. Now the condition $a_{T+1} \leq \gamma(\mu)$ together with the forward invariance of P_μ implies that $a_t \leq \gamma(\mu)$ for all $t > T$.

In summary, $a_t \leq \max\{\beta(a_0, t), \gamma(\mu)\}$ for all t, or

$$W(x(t, x^0, u)) \leq \max\{\beta(W(x^0), t), \gamma(\|u\|_\infty)\}, \; t = 0, 1, \ldots .$$

Let $\theta_i \in \mathcal{K}_\infty$ be such that $\omega(x) \leq \theta_1(W(x))$ and $W(x) \leq \theta_2(\omega(x))$. Then

$$\omega(x(t, x^0, u)) \leq \max\{\theta_1(\beta(\theta_2(\omega(x^0)), t)), \gamma(\|u\|_\infty)\}$$
$$\leq \widetilde{\beta}(\omega(x^0), t) + \gamma(\|u\|_\infty)$$

with $\widetilde{\beta}(r, t) = \theta_1(\beta(\theta_2(r), t))$ is an ISS estimate. ∎

5.5. Application to steepest descent

For each $r \geq 0$, we let

$$L(r) = \text{a Lipschitz constant for } \nabla V \text{ on the set } \{x \mid V(x) \leq r\}.$$

Without loss of generality we may take L as a continuous, nondecreasing, and everywhere nonzero function. Letting

$$\theta(r) := \frac{1}{18L(r)}$$

we conclude that:

$$|\nabla V(x)| \geq 2K_B |u| \quad \Rightarrow \quad \Delta V(x, u) \leq -\theta(V(x)) |\nabla V(x)|^2 . \quad (13)$$

For $x \in \mathcal{A}$, this is immediate since both sides vanish; for $x \notin \mathcal{A}$ it follows from (11).

From now on, we assume that:

[SG] $|\nabla V|$ is a size function for $(\mathbb{X}, \mathcal{A})$.

This property implies **[PD]**.

Theorem 4. *Suppose that* **[SV], [LL], [SG]** *hold. The system $x^+ = f(x, u)$ is ISS, and V is an ISS-Lyapunov function for it.*

Proof. We need to obtain an estimate as in (12). Let θ be as in (13). Since $|\nabla V|/(2K_B)$ is a size function, we may pick χ as any $\mathcal{K}_\infty$ function with the property that $\chi(|\nabla V(x)|/(2K_B)) \geq \omega(x)$ for all x. Now, if the pair (x, u) is such that $\omega(x) \geq \chi(|u|)$, then $\chi(|\nabla V(x)|/(2K_B)) \geq \omega(x) \geq \chi(|u|)$, and therefore $|\nabla V(x)| > 2K_B |u|$. Thus we have the implication

$$\omega(x) \geq \chi(|u|) \quad \Rightarrow \quad \Delta V(x, u) \leq -\theta(V(x)) |\nabla V(x)|^2 .$$

Since both $|\nabla V|^2$ and $V - \overline{V}$ are size functions, there is some $\widetilde{\alpha} \in \mathcal{K}_\infty$ such that $|\nabla V(x)|^2 \geq \widetilde{\alpha}(V(x) - \overline{V})$, from which

$$-\theta(V(x)) |\nabla V(x)|^2 \leq -\theta(V(x)) \widetilde{\alpha}(V(x) - \overline{V})$$

for all $x \in \mathbb{X}$. Let

$$\alpha(r) := \theta(r + \overline{V}) \widetilde{\alpha}(r).$$

Since θ is continuous and everywhere positive, and $\widetilde{\alpha} \in \mathcal{K}_\infty$, it follows that α is continuous and positive definite. We have

$$\omega(x) \geq \chi(|u|) \quad \Rightarrow \quad \Delta V(x, u) \leq -\alpha(V(x) - \overline{V})$$

and therefore V is a discrete-time ISS Lyapunov function, as claimed. ∎

We remark that in the special case that ∇V is globally Lipschitz, one can take L, and hence also θ, as a constant, so that α can be picked of class $\mathcal{K}_\infty$.

6. Discussion

We have analyzed the ISS properties of continuous-time gradient descent on open subsets of Euclidean space, as well as the ISS properties of the associated discrete-time steepest descent algorithm.

The conditions that we impose, which generalize the Polyak–Łojasiewicz condition, have appeared in the recent literature in similar contexts. For example, in [28] one finds extremum-seeking controllers based on gradient flows and an ISS property with respect for disturbances, for an integrator and a kinematic unicycle; $\mathbb{X}$ is a closed submanifold of $\mathbb{R}^n$. The paper [29] studies the gradient minimization of a function $V_q(x)$ on $\mathbb{X} = \mathbb{R}^n$, where a parameter q represents time-varying uncertainty and an ISS property is established with respect to the rate of change of q (which is a notion called "DISS" in [30]). The work [31] solves an output regulation problem for switched linear dynamical systems, with $\mathbb{X} = \mathbb{R}^n$, proving an ISS property for gradient flows with respect to unknown disturbances acting on the plant. In [32], the authors also study a gradient flow and show ISS with respect to additive errors, assuming strong convexity of the function to be minimized (in fact, a more general "convex-concave" property), also with $\mathbb{X} = \mathbb{R}^n$.

It is worth remembering that ISS theory provides an overall conceptual view, and is never the whole story. To be useful in specific applications, good estimates of the various gain functions are required. An analogy is Lyapunov-function analysis of nonlinear differential equations: while showing stability is an important first step, in practice one wants quantifications of overshoots, speed of convergence, and so on. The brief discussion of the LQR problem emphasizes that most of the actual work goes into establishing such estimates, as is the case in the various works that we have cited. Nonetheless, it seems useful to have a conceptual framework and "roadmap" that helps organize the overall abstract ideas.

Even at the conceptual level, there are many extensions still to be explored. We mentioned the extension to Riemannian manifolds. Also, finite- and fixed-time gradient flows [33,34] may be studied on open subsets and, most importantly in the current context, from the point of view of finite and fixed-time ISS in the sense of e.g. [35–37]; a recent contribution along those lines is [38].

CRediT authorship contribution statement

Eduardo D. Sontag: Conceptualization, Methodology, Writing – original draft, Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author thanks Yuan Wang and Denis Efimov for a critical reading and suggestions, the reviewers for many constructive suggestions, and the Associate Editor for the idea of Remark 3.4. This research was supported in part by grants ONR, United States N00014-21-1-2431 and AFOSR, United States FA9550-21-1-0289.

References

[1] M. Hesameddin, A. Zare, M. Soltanolkotabi, M.R. Jovanović, Convergence and sample complexity of gradient methods for the model-free linear quadratic regulator problem, 2021, arXiv:1912.11899.

[2] L. Grüne, Input-to-state stability, numerical dynamics and sampled-data control, GAMM-Mitt. 31 (1) (2008) 94–114.

[3] E.D. Sontag, Smooth stabilization implies coprime factorization, IEEE Trans. Automat. Control 34 (4) (1989) 435–443.

[4] E.D. Sontag, Input to state stability: Basic concepts and results, in: P. Nistri, G. Stefani (Eds.), Nonlinear And Optimal Control Theory, Springer-Verlag, Berlin, 2007, pp. 163–220.

[5] E.D. Sontag, Mathematical Control Theory. Deterministic Finite-DImensional Systems, second ed., in: Texts in Applied Mathematics, vol. 6, Springer-Verlag, New York, 1998.

[6] H. Khalil, Nonlinear Systems, third ed., Prentice Hall, Upper Saddle River, NJ, 2002.

[7] A. Isidori, Nonlinear Control Systems II, Springer-Verlag, London, 1999.

[8] J. Kurzweil, On the inversion of Ljapunov's second theorem on stability of motion, Czechoslovak Math. J. 6 (1956) 217–259, English translation in American Mathematical Society Translations (2), 24, 19-77..

[9] N.N. Krasovskii, Stability of Motion, Stanford University Press, Stanford, CA, 1963.

[10] V. Lakshmikantham, S. Leela, Differential And Integral Inequalities, Academic Press, 1969.

[11] A.R. Teel, L. Praly, A smooth Lyapunov function from a class-$\mathcal{KL}$ estimate involving two positive semidefinite functions, Control, Optim. Calc. Var. 5 (2000) 313–367.

[12] N. Noroozi, A. Khayatian, R. Geiselhart, A characterization of integral input-to-state stability for hybrid systems, Math. Control, Signals, Syst. 29 (3) (2017) 13.

[13] C. Cai, A.R. Teel, Robust input-to-state stability for hybrid systems, SIAM J. Control Optim. 51 (2) (2013) 1651–1678.

[14] D. Angeli, D. Efimov, Characterizations of input-to-state stability for systems with multiple invariant sets, IEEE Trans. Autom. Control 60 (12) (2015) 3242–3256.

[15] W. Hahn, Stability of Motion, Springer-Verlag, Berlin, 1967.

[16] E.D. Sontag, Y. Wang, Notions of input to output stability, Systems Control Lett. 38 (4–5) (1999) 235–248.

[17] Y. Lin, E.D. Sontag, Y. Wang, A smooth converse Lyapunov theorem for robust stability, SIAM J. Control Optim. 34 (1) (1996) 124–160.

[18] E.D. Sontag, Y. Wang, On characterizations of the input-to-state stability property, Systems Control Lett. 24 (5) (1995) 351–359.

[19] F.W. Wilson, The structure of the level surfaces of a Lyapunov function, J. Differential Equations 3 (1967) 323–329.

[20] Iu E. Nesterov, Introductory lectures on convex optimization: A basic course, in: Introductory Lectures on Convex Optimization : A Basic Course, in: Mathematics and its Applications, vol. 564, Kluwer Academic Publishers, Boston, 2004-2004.

[21] H.T. Toivonen, A globally convergent algorithm for the optimal constant output feedback problem, Int. J. Control 41 (6) (1985) 1589–1599.

[22] W. Levine, M. Athans, On the determination of the optimal constant output feedback gains for linear multivariable systems, IEEE Trans. Autom. Control 15 (1) (1970) 44–48.

[23] T. Rautert, E.W. Sachs, Computational design of optimal output feedback controllers, SIAM J. Optim. 7 (3) (1997) 837–852.

[24] H. Karimi, J. Nutini, M. Schmidt, Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition, in: Paolo Frasconi, N. Landwehr, G. Manco, J. Vreeken (Eds.), Machine Learning And Knowledge Discovery in Databases, Springer International Publishing, Cham, 2016, pp. 795–811.

[25] Z-P. Jiang, E.D. Sontag, Y. Wang, Input-to-state stability for discrete-time nonlinear systems, in: Proc. 14th IFAC World Congress, Vol. E (Beijing), 1999, pp. 277–282.

[26] Z.-P. Jiang, Y. Wang, Input-to-state stability for discrete-time nonlinear systems, Automatica 37 (6) (2001) 857–869.

[27] Z.-P. Jiang, Y. Wang, A converse Lyapunov theorem for discrete-time systems with disturbances, Systems Control Lett. 45 (1) (2002) 49–58.

[28] R. Suttner, S. Dashkovskiy, Robustness properties of a large-amplitude, high-frequency extremum seeking control scheme, 2021, arXiv:2009.14676.

[29] J.I. Poveda, M. Krstić, Nonsmooth extremum seeking control with user-prescribed fixed-time convergence, IEEE Trans. Autom. Control 66 (12) (2021) 6156–6163.

[30] D. Angeli, E.D. Sontag, Y. Wang, Input-to-state stability with respect to inputs and their derivatives, Int. J. Robust Nonlinear Control 13 (11) (2003) 1035–1056.

[31] G. Bianchin, J.I. Poveda, E. Dall'Anese, Online optimization of switched LTI systems using continuous-time and hybrid accelerated gradient flows, 2020, arXiv:2008.03903.

[32] A. Cherukuri, E. Mallada, S. Low, J. Cortés, The role of convexity in saddle-point dynamics: Lyapunov function and robustness, IEEE Trans. Autom. Control 63 (8) (2018) 2449–2464.

[33] K. Garg, D. Panagou, Fixed-time stable gradient flows: Applications to continuous-time optimization, IEEE Trans. Autom. Control 66 (5) (2021) 2002–2015.

[34] O. Romero, M. Benosman, Finite-time convergence in continuous-time optimization, in: Hal Daumé III, Aarti Singh (Eds.), Proceedings of the 37th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 119, PMLR, 2020, pp. 8200–8209.

[35] Y. Hong, Z.-P. Jiang, G. Feng, Finite-time input-to-state stability and applications to finite-time control, IFAC Proc. Vol. 41 (2) (2008) 2466–2471.

[36] F. Lopez-Ramirez, D. Efimov, A. Polyakov, W. Perruquetti, Finite-time and fixed-time input-to-state stability: Explicit and implicit approaches, Systems Control Lett. 144 (2020) 104775.

[37] J. Holloway, M. Krstic, Prescribed-time output feedback for linear systems in controllable canonical form, Automatica 107 (2019) 77–85.

[38] J.I. Poveda, M. Krstic, Fixed-time seeking and tracking of time-varying extrema, in: 2021 IEEE Conference on Decision And Control, CDC, 2021, in press.