STABILITY OF INTERCONNECTED SECTOR-BOUNDED SYSTEMS, WITH APPLICATION TO DESIGNING OPTIMIZATION ALGORITHMS

by

Saman Cyrus

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Department of Electrical & Computer Engineering)

at the

UNIVERSITY OF WISCONSIN–MADISON

2021

Date of final oral examination: 01/08/2021

The dissertation is approved by the following members of the Final Oral Committee:
Laurent Lessard, Associate Professor, Electrical and Computer Engineering
Victor M. Zavala, Associate Professor, Chemical and Biological Engineering
William A. Sethares, Professor, Electrical and Computer Engineering
Xiangru Xu, Assistant Professor, Mechanical Engineering

To my parents, Simin Moslehi Mosleh Abadi and Homayoun (AhmadAli) Sirous
and my brothers Ashkan and Peyman

ACKNOWLEDGMENTS

First and foremost, I would like to thank my adviser Professor Laurent Lessard. If it wasn't for Laurent's constant support and help, finishing this PhD would've been impossible. All these years, not only Laurent has been a great teacher and a spectacular mentor for me, he has also been my friend who I really appreciate in my life.

I would also thank my committee members, Professor Victor Zavala, William Sethares, and Xiangru Xu for their useful comments and feedback.

Last but not least, I want to thank my dear family who are always in my heart. My mother and father, Simin Moslehi Mosleh-Abadi and Ahmad-Ali Sirous, and my brothers Ashkan and Peyman. I appreciate your patience and constant support all these years and I am really grateful for your help.

I would like to thank all people who supported me during my graduate studies at UW-Madison. Without your help, the journey would've been much longer and less fun.

DISCARD THIS PAGE

TABLE OF CONTENTS

APPENDICES

DISCARD THIS PAGE

LIST OF TABLES

DISCARD THIS PAGE

LIST OF FIGURES

List of Symbols

$\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$	Integers, real numbers, and complex numbers respectively
$\mathbb{Z}_+$, $\mathbb{R}_+$	Non-negative integers and real numbers
$\mathbb{F}$	The field of either real numbers $\mathbb{R}$, or complex numbers $\mathbb{C}$
$\bar{x}$	Complex conjugate of $x \in \mathbb{F}$
X^*	Conjugate transpose of a matrix $X \in \mathbb{F}^{m,n}$
$\preceq$, $\prec$, $\succ$, $\succeq$	(semi) definite partial ordering of matrices in $\mathbb{F}^{n \times n}$
$\mathcal{V}$	A semi-inner product space
$\Vert \cdot \Vert$	General norm
$\langle \cdot, \cdot \rangle$	Semi-inner product
$f(t)$	Time-domain continuous-time scalar signal
$\hat{f}(\omega)$	Fourier transform of signal $f(t) \in L_2$, defined as $\hat{f}(\omega) := \int_{-\infty}^{\infty} e^{-j\omega t} f(t) \, \mathrm{d}t$
$\hat{f}(\omega)$	Discrete-time Fourier transform of signal $f(t) \in \ell_2$, defined as $\hat{f}(\omega) := \sum\limits_{t=-\infty}^{\infty} e^{-j\omega t} f(t)$
R	A relation R on $\mathcal{V}$
$\mathscr{R}(\mathcal{V})$	The set of all relations on $\mathcal{V}$
$\mathrm{dom}(R)$	The domain of R
$\hat{G}(s)$	Transfer matrix of continuous-time system G
$\tilde{G}(z)$	Pulse transfer matrix of discrete-time system G
$\langle x, y \rangle_T^\rho$	The family of semi-inner products as $\sum_{k=0}^{T} \rho^{-2k} \langle x[k], y[k] \rangle$
$\mathcal{L}_2$	A Lebesgue space of functions $\mathcal{T} \to \mathbb{F}^n$
$\ell_2^n(\mathbb{F})$	The space of square-summable sequences with $\langle x, y \rangle := \sum_{t=0}^{\infty} x(t)^* y(t)$
$L_2^n(\mathbb{F})$	The space of square-integrable functions with $\langle x, y \rangle := \int_0^{\infty} x(t)^* y(t) \, \mathrm{d}t$
$x_T(t)$	Truncation at time T for signal $x : \mathcal{T} \to \mathbb{F}^n$
$\mathcal{L}_{2\mathrm{e}}$	The extended spaces $\mathcal{L}_{2\mathrm{e}} \supseteq \mathcal{L}_2$, defined as $\mathcal{L}_{2\mathrm{e}} := \{x : \mathcal{T} \to \mathbb{F}^n \mid x_T \in \mathcal{L}_2 \text{ for all } T \geq 0\}$

$\mathcal{H}$	Hilbert space
$\mathscr{L}(\mathcal{V})$	The set of all linear relations
$\mathcal{C}_G$, $\mathcal{C}_\Phi$	Arbitrary sets of (possibly nonlinear) relations
$\mathscr{F}(\mathcal{L}_{2\mathrm{e}})$	The set of causal operators on $\mathcal{L}_{2\mathrm{e}}$
$\mathscr{L}(\mathcal{L}_{2\mathrm{e}})$	The set of causal linear operators on $\mathcal{L}_{2\mathrm{e}}$
$\mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2\mathrm{e}})$	The set of linear time-invariant (LTI) operators on $\mathcal{L}_{2\mathrm{e}}$
$\mathbf{G}(s)$	The transfer function of system G
$\hat{G}(\omega)$	The frequency response of G defined as $\hat{G}(\omega) = \mathbf{G}(j\omega)$ when $\mathcal{L}_{2\mathrm{e}} = L_{2\mathrm{e}}$
$\hat{G}(\omega)$	The frequency response of G defined as $\hat{G}(\omega) = \mathbf{G}(e^{j\omega})$ for $\mathcal{L}_{2\mathrm{e}} = \ell_{2\mathrm{e}}$

ABSTRACT

Stability of interconnected systems has been a point of interest for decades. Historically, two parallel paths have been pursued: stability results based on Lyapunov's work in the former Soviet union (see for example [1]) vs. input-output stability in the west [2] using functional analysis. The idea is to obtain conditions that guarantee robust stability of the feedback interconnection of two given systems, either against a given set of inputs, or against a given set of initial conditions. For linear systems, these two notions of stability, in many cases, lead to identical results, i.e., there aren't many different notions for stability. For nonlinear systems on the other hand, changing the flavor of stability which is the point of interest, can change the required conditions (look at [3, §6.3] to see how Lyapunov-style results and input-output style results are related). In addition, while stability criteria for linear time-invariant systems are necessary and sufficient, available tools for stability analysis of nonlinear feedback systems, are sufficient-only and not *necessary and sufficient* in the general case.

In this dissertation, we first develop a robust boundedness theorem defined over a general semi-inner product space (Theorem 3.1). We describe the required conditions for which this theorem is both necessary and sufficient for robust boundedness. To motivate application of this theorem into real systems, we obtain versions of this theorem which include time-domain signals and therefore provide sufficient conditions for robust stability (Corollary 3.1). We show when these results are both necessary and sufficient (Theorem 3.4). Then we show how tools that are developed in the literature for analysis and synthesis of robust stability of interconnected systems can be used to find upper-bounds for convergence of first-order optimization algorithms, applied to the family of smooth strongly convex functions.

Chapter 2 introduces required mathematical tools for this dissertation. In Chapter 3, we develop a unifying necessary and sufficient condition to guarantee stability of a class of nonlinear feedback systems. The main theorem of Chapter 3 (Theorem 3.1) is developed for a semi-inner product vector space and therefore can be specialized to other vector spaces, i.e., Lebesgue spaces (ℓ_{2e}, L_{2e} etc.), weighted Lebesgue spaces (ℓ_{2e}^{ρ}, etc.), Euclidean space ($\mathbb{R}^n$), and etc. In Chapter 4, first we write the problem of first-order optimization of strongly convex functions as a feedback interconnection of two systems, and then using the tools that are available in the literature of robust control, i.e., Zames-Falb multipliers and circle criterion, we develop a fast robust to (multiplicative) noise algorithm for this class of systems. Using numerical simulations

we show that the suggested algorithm, on one extreme, is as fast as the current fastest first-order optimization algorithm in the literature (Triple momentum method), and on the other extreme, it is as robust to noise as the Gradient Method (GM) with step-size $\alpha = 1/L$ which is the most robust algorithm currently in the literature. Chapter 5 suggests future directions to continue this line of research.

Chapter 1

Introduction

Robust stability of feedback systems has been a point of interest for decades in the literature of control theory. the origin of this problem goes back to the theories of “absolute stability”[1] and “input-output” stability [5], as well as dissipative systems’ theory [6, 7]. Later the idea of input-output stability involved using multipliers to decrease conservativity of the design. The theory of multipliers, originally proposed by O’shea [8, 9] and later formalized by Zames and Falb [2].

For a forced feedback interconnection (Figure 1.1), finding the required conditions for input-output stability of the interconnection is a point of interest while for an unforced feedback interconnection (Figure 1.2), global asymptotic stability of the equilibrium point is the point of interest. In the case that one of the systems is a linear time-invariant (LTI) system, the problem of ensuring the stability of an unforced feedback system is called “the Lur’e problem”.

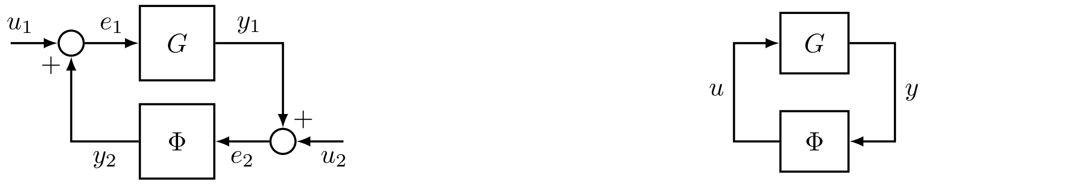


Figure 1.1: Forced feedback interconnection.

Figure 1.2: Unforced feedback interconnection.

Figure 1.3: Forced vs. unforced feedback interconnections.

Classical tools to ensure robust stability of a feedback interconnection includes the circle criterion [10], small-gain theorem [5], passivity theorem [11], conic systems’ theorem [12], and Integral Quadratic Constraints (IQCs) [13]. Most of these methods, fix conditions on one of the systems of the feedback interconnection and then find the required conditions on the other system that guarantees stability and it has been

[1] The feedback interconnection of a linear system with a static nonlinearity is called absolutely stable if the closed-loop system is stable for every nonlinearity in a given class [4].

shown in the literature that these theorems to investigate stability of a system, under certain circumstances, are equivalent[2] [15].

One of the most famous of these classical theorems is the circle criterion, which has been used to evaluate both input-output stability of a forced feedback interconnection and global exponential stability of an unforced feedback interconnection, i.e., absolute stability of the Lur'e problem (see [1,3] for continuous-time single-input single-output absolute stability and [16–20] for discrete-time single-input single-output absolute stability). It has been shown [10,21,22] that, in general, the circle criterion is a *sufficient* but *not necessary* condition for absolute stability. There have been efforts to clarify this gap between guaranteed stability and guaranteed instability [23–28], therefore, there is an intermediate case where the circle criterion cannot guarantee either stability or instability.

In Chapter §3, we suggest a result in the spirit of the aforementioned classical results but expressed in a semi-inner product space which generalizes the classical results while proposes necessary and sufficient conditions for stability of a certain class of feedback systems.

Another point of interest of this dissertation is the application of stability tools, developed for feedback interconnected systems, to designing first-order optimization algorithms. Recently, feedback interconnections and their stability have been used in the optimization literature [29–31]. In [32], it is shown that first-order optimization algorithms, i.e., Gradient Method (GM), Nesterov's Fast Gradient Method (FGM) [33], Heavy-ball Method (HB) [34] and Triple Momentum Method (TMM) [11] can be formulated as a Lur'e problem. This makes the problem of finding the sufficient condition for absolute stability of the feedback interconnection equivalent to certifying an upper-bound for the linear convergence of first-order optimization algorithms. Using Integral Quadratic Constraints (IQC), a less conservative upper-bound for the convergence rate of these first-order algorithms was found in [32]. In this document, we use design tools from the robust control literature to design a robust-to-noise first-order optimization algorithm; called Robust Momentum Method (RMM). Section 4, investigates the behavior of RMM in comparison to other famous algorithms in the literature using simulations. We characterize the performance of RMM under multiplicative gradient noise and show how to tune RMM to trade off convergence rate and robustness to noise.

[2] In the sense that stability of system A_1 studied with method M_1 is equivalent to stability of the transformed system A_2, assessed with method M_2, where system A_1 has been transformed to A_2, using an appropriate loop-transformation. Note that this does not mean that applying methods M_1 and M_2 to the same system A_1 produces identical results. In [14] it has been shown that if these tools are applied to the same system, very different results will be obtained.

Chapter 2

Mathematical Preliminaries

In this chapter, we provide the definitions, theorems and lemmas that are essential for this dissertation.

2.1 Notation

Preliminaries The set $\mathbb{F}$ refers to either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$. We use $\mathbb{R}_+$ ($\mathbb{Z}_+$) to denote the nonnegative real numbers (integers). We write the complex conjugate of $x \in \mathbb{F}$ as $\bar{x}$ and the conjugate transpose of a matrix $X \in \mathbb{F}^{m\times n}$ as X^*. We use the symbols $\preceq$, $\prec$, $\succ$, $\succeq$ to denote the (semi)definite partial ordering of matrices in $\mathbb{F}^{n\times n}$. A matrix $A = A^* \in \mathbb{F}^{n\times n}$ is ***indefinite*** if $A \npreceq 0$ and $A \nsucceq 0$.

Linear time-invariant (LTI) systems with n states, m inputs, and p outputs are shown as

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{2.1a}$$

$$y(t) = Cx(t) + Du(t), \tag{2.1b}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$ and $t \in \mathbb{R}_+$. Discrete-time systems are sometimes represented by

$$x_{k+1} = Ax_k + Bu_k, \tag{2.2a}$$

$$y_k = Cx_k + Du_k. \tag{2.2b}$$

to follow the notation of optimization literature.

Semi-inner products A *semi-inner product space* is a vector space $\mathcal{V}$ over a field $\mathbb{F}$ equipped with a semi-inner product[1]$\langle \cdot, \cdot \rangle$. This is identical to an inner product except that the associated norm is a seminorm[2]. In other words, $\|x\| := \sqrt{\langle x, x\rangle} \geq 0$ for all $x \in \mathcal{V}$, but $\|x\| = 0$ need not imply that $x = 0$.

[1] We use the convention that a semi-inner product is linear in its second argument, so $\langle x, ay + bz\rangle = a\langle x, y\rangle + b\langle x, z\rangle$ for all $x, y, z \in \mathcal{V}$ and $a, b \in \mathbb{F}$. Also, $\langle x, y\rangle = \overline{\langle y, x\rangle}$.

[2] The reason for using semi-inner product spaces is that in $\mathcal{L}_{2e}$, we use this inner product $\langle \cdot, \cdot \rangle_T$, where we truncate after time T. This is not an inner product. It's a semi-inner product. Because if two signals $x(t)$ and $y(t)$ are the same on $[0, T]$ but are different for $T > 0$, then we still have $\|x - y\|_T = 0$, even though x and y are different.

Relations A *relation* R on $\mathcal{V}$ is a subset of the product space $R \subseteq \mathcal{V} \times \mathcal{V}$. We write $\mathcal{R}(\mathcal{V})$ to denote the set of all relations on $\mathcal{V}$. The ***domain*** of R is defined as $\operatorname{dom}(R) := \{x \in \mathcal{V} \mid (x, y) \in R \text{ for some } y \in \mathcal{V}\}$. For any $x \in \operatorname{dom}(R)$, we write Rx to denote any $y \in \mathcal{V}$ such that $(x, y) \in R$.

We define $\mathcal{V}^2$ to be the augmented vectors $u =: \left(\begin{smallmatrix} u_1 \\ u_2 \end{smallmatrix}\right)$ where $u_1, u_2 \in \mathcal{V}$. We overload matrix multiplication to have an intuitive interpretation in $\mathcal{V}^2$. Specifically, for any $\xi, \zeta \in \mathcal{V}^2$ and any matrix $N \in \mathbb{F}^{2\times 2}$,

$$N\xi = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} := \begin{bmatrix} N_{11}\xi_1 + N_{12}\xi_2 \\ N_{21}\xi_1 + N_{22}\xi_2 \end{bmatrix} \in \mathcal{V}^2.$$

Likewise, inner products in $\mathcal{V}^2$ have the interpretation

$$\langle \xi, \zeta \rangle = \left\langle \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \right\rangle := \langle \xi_1, \zeta_1 \rangle + \langle \xi_2, \zeta_2 \rangle .$$

The closed-loop system of Fig. 3.1 defines relations:

$$R_{uy} := \left\{ (u, y) \in \mathcal{V}^2 \times \mathcal{V}^2 \mid (3.1) \text{ holds for some } e \in \mathcal{V}^2 \right\}$$
$$R_{ue} := \left\{ (u, e) \in \mathcal{V}^2 \times \mathcal{V}^2 \mid (3.1) \text{ holds for some } y \in \mathcal{V}^2 \right\}$$

We call a set of relations $\mathcal{C} \subseteq \mathcal{R}(\mathcal{V})$ ***feedback-invariant*** if it is closed under feedback interconnection. Specifically, $\{(u_i, y_j) \mid (u, y) \in R_{uy}\} \in \mathcal{C}$ for all $G, \Phi \in \mathcal{C}$ and for all $i, j \in \{1, 2\}$. We call $\mathcal{C}$ ***complete*** if given any $x, y \in \mathcal{V}$, there exists $\Phi \in \mathcal{C}$ such that $(x, y) \in \Phi$.

2.1.1 Notation for Extended Spaces

In this dissertation, $\mathcal{L}_2$ denotes a Lebesgue space of functions $\mathcal{T} \to \mathbb{F}^n$. If $\mathcal{T} = \mathbb{Z}_+$, this could be the space $\ell_2^n(\mathbb{F})$ of square-summable sequences with $\langle x, y \rangle := \sum_{t=0}^{\infty} x(t)^* y(t)$. Likewise, if $\mathcal{T} = \mathbb{R}_+$, this could be the space $L_2^n(\mathbb{F})$ of square-integrable functions with $\langle x, y \rangle := \int_0^\infty x(t)^* y(t)\, \mathrm{d}t$. The definitions that follow apply to commonly used time domains $\mathcal{T}$ such as $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{Z}$, and $\mathbb{Z}_+$.

For any $T \geq 0$ and function $x : \mathcal{T} \to \mathbb{F}^n$, the ***truncation*** at time T is the function $x_T : \mathcal{T} \to \mathbb{F}^n$ defined as

$$x_T(t) := \begin{cases} x(t) & t \leq T \\ 0 & t > T. \end{cases}$$

The ***extended spaces*** $\mathcal{L}_{2\mathrm{e}} \supseteq \mathcal{L}_2$ are defined as

$$\mathcal{L}_{2\mathrm{e}} := \{ x : \mathcal{T} \to \mathbb{F}^n \mid x_T \in \mathcal{L}_2 \text{ for all } T \geq 0 \} .$$

We overload the notation $\langle \cdot, \cdot \rangle_T$ to denote the truncated semi-inner product on $\mathcal{L}_{2\mathrm{e}}$. That is, $\langle x, y \rangle_T := \langle x_T, y_T \rangle$. We also define the associated seminorm $\|x\|_T := \sqrt{\langle x, x \rangle_T}$.

Definition 2.1 (causal operators)**.** *A **causal** operator on $\mathcal{L}_{2e}$ is a function $G : \mathcal{L}_{2e} \to \mathcal{L}_{2e}$ with the property that for all $f \in \mathcal{L}_{2e}$ and $T \geq 0$, we have $(Gf)_T = (Gf_T)_T$.*

The *shift operator* $S_\tau : \mathcal{L}_{2\mathrm{e}} \to \mathcal{L}_{2\mathrm{e}}$ is defined as $(S_\tau f)(t) = f(t-\tau)$.[3] A causal operator G on $\mathcal{L}_{2\mathrm{e}}$ is said to be ***time-invariant*** if $GS_\tau f = S_\tau Gf$ for all $f \in \mathcal{L}_{2\mathrm{e}}$ and all τ. We denote the set of ***linear time-invariant*** (LTI) operators on $\mathcal{L}_{2\mathrm{e}}$ as $\mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2\mathrm{e}})$.

Given $G \in \mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2\mathrm{e}})$, we let $\hat{G} : \mathbb{R} \to \mathbb{C}$ denote the ***frequency response*** of G, defined as follows. Let G have transfer function $\mathbf{G}$. If $\mathcal{L}_{2\mathrm{e}} = L_{2\mathrm{e}}$, then $\hat{G}(\omega) := \mathbf{G}(j\omega)$. If $\mathcal{L}_{2\mathrm{e}} = \ell_{2\mathrm{e}}$, then $\hat{G}(\omega) := \mathbf{G}(e^{j\omega})$. We write $\mathcal{L}_\infty$ to denote the set of frequency responses that are essentially bounded for all $\omega \in \mathbb{R}$.

Definition 2.2 (well-posedness)**.** *If $G, \Phi \in \mathscr{F}(\mathcal{L}_{2e})$, the interconnection of Fig. 3.1 is said to be **well-posed** if $R_{uy} \in \mathscr{F}(\mathcal{L}_{2e})$ (equivalently, $R_{ue} \in \mathscr{F}(\mathcal{L}_{2e})$). In other words, the closed-loop relation should be a causal operator on $\mathcal{L}_{2e}$. In this case, we will refer to R_{uy} as the* closed-loop map*.*

Family of functions of interest in optimization chapter: The set of functions that are m-strongly convex and L-smooth is denoted $\mathcal{F}(m, L)$. In particular, $f \in \mathcal{F}(m, L)$ if for all $x, y \in \mathbb{R}^n$,

$$m\|x-y\|^2 \leq (\nabla f(x) - \nabla f(y))^{\mathsf{T}} (x-y) \leq L\|x-y\|^2.$$

The condition ratio is defined as $\kappa := L/m$.

2.2 Norms

The notion of norm is used extensively in this document. Therefore, it is necessary to define what we mean by norm.

Definition 2.3 (vector norms [5])**.** *Let $v \in \mathbb{C}^n$ and defined as $v = (v_1, v_2, \ldots)$, then we can define*

$$\|v\|_p \triangleq \begin{cases} \left(\sum\limits_{i=1}^{n} |v_i|^p\right)^{1/p}, & p \in [1, \infty) \\ \max\limits_i |v_i|, & p = \infty \end{cases}$$

Definition 2.4 (Matrix Norms [5])**.** *For a matrix $A = [a_{ij}]$ defined on the set of all $n \times n$ matrices, norm is defined as*

$$\|A\|_p \triangleq \begin{cases} \max\limits_j \sum\limits_{i=1}^{n} |a_{ij}|, & p = 1 \\ \max\limits_i [\lambda_i(A^*A)]^{1/2}, & p = 2 \\ \max\limits_i \sum\limits_{j=1}^{n} |a_{ij}|, & p = \infty \end{cases}$$

where $\lambda_i(M)$ is the ith eigenvalue of M.

[3] By convention, if $\mathcal{L}_{2\mathrm{e}}$ is defined on the time interval $[0, \infty)$, then $f(t) = 0$ for all $t < 0$. So $S_\tau f$ is well-defined for all $f \in \mathcal{L}_{2\mathrm{e}}$ and all τ.

Definition 2.5 (Vector-valued functions Norms [5]). *Let*

$$E = \{f : \mathbb{R} \to \mathbb{R}^n, f = (f_1, f_2, \cdots, f_n) \mid \textit{all } f_i\textit{'s locally (Lebesgue) integrable}\}$$

Then we define

$$\|f\|_p \triangleq \left(\int \|f(t)\|^p \mathrm{d}t \right)^{1/p}, \qquad p \in [1, \infty)$$

$$\|f\|_\infty \triangleq \textit{ess} \sup_{t \in \mathbb{R}} \|f(t)\|,$$

where $\|f(t)\|$ is a vector-valued norm at time t, i.e., a norm on $\mathbb{R}^n$.

2.3 Sector-Bounded systems:

Definition 2.6 (Sector-bounded memoryless but possibly time-varying nonlinearity). *Suppose $\phi : \mathbb{F}_+ \times \mathbb{R} \to \mathbb{R}$ is a given function, and define a corresponding operator Φ on $\mathcal{L}_{1e}$ by*

$$(\Phi x)(t) = \phi(t, x(t)), \quad \forall t \geq 0 \tag{2.4}$$

We say that ϕ or Φ belongs to the sector $[a, b]$ if

$$\phi(t, 0) = 0, \quad a \leq \frac{\phi(t, x)}{x} \leq b, \quad \forall x \neq 0, \quad \forall t \geq 0 \tag{2.5}$$

Definition 2.7. *Consider a causal but not necessarily memoryless function $\Phi : \mathcal{L}_{2e}^m \to \mathcal{L}_{2e}^l$ and assume that $a < b$. Then we say that Φ belongs to the sector $[a, b]$ if*

$$\left\| \Phi x - \frac{a+b}{2} x \right\|_T \leq \frac{b-a}{2} \|x\|_T, \quad \forall T \geq 0, \quad \forall x \in \mathcal{L}_{2e}^m \tag{2.6}$$

where $\| \cdot \|_T$ is the truncated 2-seminorm.

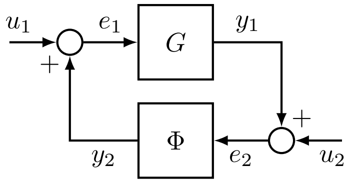
Note that by defining Φ as it is defined in (2.4), in SISO case, we obtain (2.5).

2.4 Stability

Here we review the notion of stability (both Lyapunov-flavor and Input-output flavor) for linear systems, nonlinear systems and interconnected systems. In addition, we discuss how Lyapunov-flavored results are related to input-output stability results. More in-depth literature related to stability notions that are being discussed here can be found in [3, 35–37].

2.4.1 Input-Output Stability

The definitions in this section are taken from [3].



$$e_1 = u_1 - y_2 \tag{2.7a}$$

$$e_2 = u_2 + y_1 \tag{2.7b}$$

$$y_1 = G_1 e_1 \tag{2.7c}$$

$$y_2 = G_2 e_2 \tag{2.7d}$$

Figure 2.1: An interconnected system.

$\mathcal{L}_2$ (ℓ_2)-stability of a system: A dynamical system $H : \mathcal{V}_e \to \mathcal{V}_e$ is $\mathcal{L}_2^m$ stable if $u \in \mathcal{L}_2^m \implies Hu \in \mathcal{L}_2^m$. A finite-dimensional continuous-time linear time-invariant (LTI) system $\hat{H}(s)$ is L_2^m stable if and only if all poles have negative real parts (This result is known as uniform BIBO[4] stability). Equivalently, under these conditions, the corresponding minimal state-space realization (A, B, C, D) is asymptotically stable. A discrete time LTI system $\tilde{H}(z)$ is ℓ_2^m stable if and only if all poles have magnitude less than one, this is known as uniform BIBO stability for discrete-time systems.

Finite-gain $\mathcal{L}_p^m(\ell_p^m)$ stability: Suppose R is a binary relation on $\mathcal{L}_{pe}^m$. Then R is $\mathcal{L}_p^m$-stable with finite gain if it is $\mathcal{L}_p^m$-stable and in addition there exist finite constants γ_p and b_p such that

$$(x, y) \in R, x \in \mathcal{L}_p^m \Rightarrow \|y\|_p \leq \gamma_p \|x\|_p + b_p.$$

Moreover, R is $\mathcal{L}_p$-stable with finite gain and zero bias if it is $\mathcal{L}_p^m$-stable and in addition there exists a finite constant γ_p such that

$$(x, y) \in R, x \in \mathcal{L}_p^m \Rightarrow \|y\|_p \leq \gamma_p \|x\|_p.$$

In this dissertation, the notion of interest of *stability* for LTI systems is stability with finite gain and zero bias, which we simply call stability.

Remark 2.1. *Suppose $A : \mathcal{L}_{pe}^n \to \mathcal{L}_{pe}^m$ is causal and $\mathcal{L}_p$-stable with finite gain, and choose constants γ_p and b_p such that*

$$\|Ax\|_p \leq \gamma_p \|x\|_p + b_p, \quad \forall x \in \mathcal{L}_p^n$$

Then

$$\|Ax\|_{Tp} \geq \gamma_p \|x\|_{Tp} + b_p, \quad \forall T \geq 0, \forall x \in \mathcal{L}_{pe}^n.$$

Interconnected systems: For a feedback system described by

where e_1, e_2, y_1, y_2, u_1, u_2 belong to the extended functional space $\mathcal{L}_{2e} = \mathcal{L}_{2e}[0, \infty)$, and G_1 , G_2 map $\mathcal{L}_{2e}$ into itself. It is assumed that for each $u_1, u_2 \in \mathcal{L}_2$ there exist $e_1, e_2, y_1, y_2 \in \mathcal{L}_{2e}$ such that (2.7) hold.

[4] Bounded-input bounded-output

The system (2.7) is said to be stable if whenever $u_1, u_2 \in \mathcal{L}_2$ any corresponding $e_1, e_2, y_1, y_2 \in \mathcal{L}_{2e}$ such that (2.7) hold, actually belong to $\mathcal{L}_2$. The system (2.7) is said to be unstable otherwise.

Stability of Conic systems:

Remark 2.2. *All interior conic systems are input-output stable.*

Proof. *To see this, define $c = max(|a|, |b|)$. Then $G \in cone[a, b] \subseteq cone[-c, c]$ implies that $\|G\| \leq c$, so it is I/O stable.* ■

Remark 2.3. *All stable systems are interior conic.*

Proof. *By definition, stable systems have finite gain. Then from $\|G\| \leq c$, we can write $G \in cone[-c, c]$.* ■

Remark 2.4. *Exterior conic systems may be stable or unstable.*

Proof. Exterior conic systems can include the case where $u = 0$ generates $y \to \infty$, which by definition, is an unstable system. ■

2.5 Integral Quadratic Constraints

One unifying framework to analyze feedback interconnections of Figures 1.1 and1.2 is integral quadratic constraints (IQCs) [13,32]. An IQC in continuous-time can be defined as:

Definition 2.8 (Continuous-time frequency-domain IQCs [13])**.** *Two signals $v \in L_2^l$ and $w \in L_2^m$ are said to satisfy the IQC defined by Π if*

$$\int_{-\infty}^{\infty} \begin{bmatrix} \hat{v}(j\omega) \\ \hat{w}(j\omega) \end{bmatrix}^* \Pi(j\omega) \begin{bmatrix} \hat{v}(j\omega) \\ \hat{w}(j\omega) \end{bmatrix} \mathrm{d}w \geq 0 \tag{2.8}$$

If (2.8) holds for all $w = \Phi v$, where Φ is a bounded operator defined as $\Phi : L_{2e}^l \to L_{2e}^m$ and $v \in L_2^l$, then we say Φ satisfies the frequency-domain continuous-time IQC defined by (2.8).

Definition 2.9 (Continuous-time time-domain IQCs [39])**.** *A causal operator $\Phi : L_{2e}^l \to L_{2e}^m$ satisfies the time-domain IQC defined by (Ψ, M) if for all $v \in \mathcal{L}_{2e}^l$, we have that $w = \Phi v$ and*

$$\int_0^T z^{\mathsf{T}}(t) M z(t) \mathrm{d}t \geq 0, \qquad T \geq 0$$

where $M = M^{\mathsf{T}} \in \mathbb{R}^{n \times n}$ and Ψ is an LTI system and z is the output of Ψ driven by inputs (v, w) with zero initial conditions.

Definition 2.10 (Discrete-time frequency-domain IQCs [40]). *Two signals $v \in \ell_2^l$ and $w \in \ell_2^m$ are said to satisfy the IQC defined by Π if*

$$\int_0^{2\pi} \begin{bmatrix} \tilde{v}(e^{j\theta}) \\ \tilde{w}(e^{j\theta}) \end{bmatrix}^* \Pi(e^{j\theta}) \begin{bmatrix} \tilde{v}(e^{j\theta}) \\ \tilde{w}(e^{j\theta}) \end{bmatrix} \mathrm{d}\theta \geq 0 \tag{2.9}$$

If (2.9) holds for all $w = \Phi v$, where Φ is a bounded operator defined as $\Phi : \ell_{2e}^l \to \ell_{2e}^m$ and $v \in \ell_2^l$, then we say Φ satisfies the frequency-domain discrete-time IQC defined by (2.9).

The following definitions include Hard and Soft time-domain IQCs for discrete-time systems [32].

Definition 2.11 (Discrete-time time-domain Hard-IQCs). *A causal operator $\Phi : \ell_{2e}^l \to \ell_{2e}^m$ satisfies the time-domain Hard-IQC defined by (Ψ, M) if for all $v \in \ell_{2e}^l$, we have that $w = \Phi v$ and*

$$\sum_{k=0}^{N} z_k^\mathsf{T} M z_k \geq 0, \qquad N \geq 0$$

where $M = M^\mathsf{T} \in \mathbb{R}^{n \times n}$ and Ψ is an LTI system and z is the output of Ψ driven by inputs (v, w) with zero initial conditions.

Definition 2.12 (Discrete-time time-domain Soft-IQCs). *A bounded, causal operator $\Phi : \ell_{2e}^l \to \ell_{2e}^m$ satisfies the time-domain Soft-IQC defined by (Ψ, M) if for all $v \in \ell_{2e}^l$, we have that $w = \Phi v$ and*

$$\sum_{k=0}^{\infty} z_k^\mathsf{T} M z_k \geq 0,$$

where $M = M^\mathsf{T} \in \mathbb{R}^{n \times n}$ and Ψ is an LTI system and z is the output of Ψ driven by inputs (v, w) with zero initial conditions.

2.6 Sector-Bounded Systems, Slope-Restricted Systems

2.6.1 Sector-Bounded Static Systems

For a good reference for the material of this section, see [5].

Definition 2.13 (SISO sector-bounded static nonlinearities). *Let $\phi : \mathbb{R} \to \mathbb{R}$ with $\phi(0) = 0$. We say that the function ϕ belongs to the sector $[a, b]$ if and only if*

$$ax^2 \leq x\phi(x) \leq bx^2, \quad \forall x \in \mathbb{R}, \quad x \neq 0$$

There are equivalent forms for the sector condition. Lemma 2.1 reviews these equivalent forms.

Lemma 2.1 (Single-Input-Single-Output equivalent forms of sector-bounded static systems). *If $\phi(0) = 0$ and $a, b \in \mathbb{R}$ and $a \leq b$, then these forms are equivalent:*

1. *For all $x \neq 0$, $a \leq \frac{\phi(x)}{x} \leq b$.*

2. *For all $x \neq 0$, $ax^2 \leq x\phi(x) \leq bx^2$.*

3. *For all $x \in \mathbb{R}$,*
$$[\phi(x) - ax][\phi(x) - bx] \leq 0.$$

4. *For all $x \in \mathbb{R}$,*
$$\left|\phi(x) - \frac{a+b}{2}x\right|^2 \leq \left(\frac{b-a}{2}\right)^2 |x|^2.$$

5. *for all $x \in \mathbb{R}$*
$$\begin{bmatrix} x \\ \phi(x) \end{bmatrix}^\mathsf{T} \begin{bmatrix} -2ab & a+b \\ a+b & -2 \end{bmatrix} \begin{bmatrix} x \\ \phi(x) \end{bmatrix} \geq 0.$$

Proof.

$1 \iff 2$**:** If we multiply item 1 by x^2, we will get item 2. Divide item 2 by x^2 and we will get item 1. Note that in the case of $x = 0$, item 2 always holds.

$3 \iff 4$ Simply

$$\begin{aligned}
[\phi(x) - ax][\phi(x) - bx] &= \phi(x)^2 - 2\frac{a+b}{2}x\phi(x) + abx^2 \\
&= (\phi(x) - \frac{a+b}{2}x)^2 - \frac{(b-a)^2}{4}x^2 \\
&= \left[\phi(x) - \frac{a+b}{2}x\right]^2 \leq \left(\frac{b-a}{2}\right)^2 x^2 \leq 0
\end{aligned}$$

$2 \implies 3$ Obviously $x\phi(x) - bx^2 \leq 0$ and $x\phi(x) - ax^2 \geq 0$, therefore $[\phi(x) - ax][\phi(x) - bx]x^2 \leq 0$.

$3 \implies 2$ Multiply both sides of 3 by x^2 and since $bx^2 \geq ax^2$, the value $x\phi(x)$ should be intermittent between these two values.

$5 \iff 3$ For all $x \in \mathbb{R}$, we can write

$$\begin{aligned}
\begin{bmatrix} x \\ \phi(x) \end{bmatrix}^\mathsf{T} \begin{bmatrix} -2ab & a+b \\ a+b & -2 \end{bmatrix} \begin{bmatrix} x \\ \phi(x) \end{bmatrix} &= (a+b)\phi(x)x - abx^2 - \phi(x)^2 \\
&= -[\phi(x) - ax][\phi(x) - bx] \geq 0
\end{aligned}$$

■

Lemma 2.2 (Multi-input multi-output equivalent forms of sector-bounded static systems)**.** *Let $\mathcal{H}$ be a Hilbert space (most of the applications use $\mathcal{H} = \mathbb{R}^n$), and assume the real scalar product $\langle \cdot, \cdot \rangle$ with associated norm $\| \cdot \|$ and let $\Phi : \mathcal{H} \to \mathcal{H}$ and $\Phi(0) = 0$ where 0 is the zero vector in $\mathcal{H}$. Assuming a and b two real constants where $b \geq a$, then the following two statements are equivalent:*

1. *For all $x \in \mathcal{H}$*

$$\langle \Phi(x) - ax, \, \Phi(x) - bx \rangle \leq 0$$

2. *For all $x \in \mathcal{H}$*

$$\left\| \Phi(x) - \frac{a+b}{2} x \right\|^2 \leq \left(\frac{b-a}{2} \right)^2 \|x\|^2$$

3. *For all $x \in L_2^m$*

$$\int\limits_{-\infty}^{\infty} \begin{bmatrix} \hat{x}(j\omega) \\ \hat{\phi}(j\omega) \end{bmatrix}^* \begin{bmatrix} -2ab & a+b \\ a+b & -2 \end{bmatrix} \begin{bmatrix} \hat{x}(j\omega) \\ \hat{\phi}(j\omega) \end{bmatrix} \mathrm{d}\omega \geq 0$$

or for all $x \in \ell_2^m$

$$\int\limits_{-\infty}^{\infty} \begin{bmatrix} \tilde{x}(e^{j\theta}) \\ \tilde{\phi}(e^{j\theta}) \end{bmatrix}^* \begin{bmatrix} -2ab & a+b \\ a+b & -2 \end{bmatrix} \begin{bmatrix} \tilde{x}(e^{j\theta}) \\ \tilde{\phi}(e^{j\theta}) \end{bmatrix} \mathrm{d}\theta \geq 0$$

Proof.

$1 \iff 2$: The proof is taken from [5]

$$\begin{aligned} 0 &\geq \langle \Phi(x) - ax, \, \Phi(x) - bx \rangle \\ &= \langle \Phi(x), \, \Phi(x) \rangle - 2 \times \frac{a+b}{2} \langle \Phi(x), \, x \rangle + \left(\frac{(a+b)^2}{4} - \frac{(b-a)^2}{4} \right) \langle x, \, x \rangle \\ &= \left\| \Phi(x) - \frac{a+b}{2} x \right\|^2 - \frac{(b-a)^2}{4} \|x\|^2 \end{aligned}$$

$1 \iff 3$: (Continuous-time case with $x \in L_2^m$) For all $x \in L_2^m$ we have

$$\langle \Phi(x) - ax, \, \Phi(x) - bx \rangle \leq 0.$$

Therefore

$$\begin{aligned} &-2 \langle \Phi(x) - ax, \, \Phi(x) - bx \rangle \\ &= -2 \left[\int_{-\infty}^{\infty} \Phi(x)^{\mathsf{T}} \Phi(x) \mathrm{d}t - b \int_{-\infty}^{\infty} \Phi(x)^{\mathsf{T}} x \mathrm{d}t - a \int_{-\infty}^{\infty} x^{\mathsf{T}} \Phi(x) \mathrm{d}t + ab \int_{-\infty}^{\infty} x^{\mathsf{T}} x \mathrm{d}t \right] \\ &= -2ab \int_{-\infty}^{\infty} x^{\mathsf{T}} x \mathrm{d}t + (a+b) \int_{-\infty}^{\infty} x^{\mathsf{T}} \Phi(x) \mathrm{d}t + (a+b) \int_{-\infty}^{\infty} \Phi(x)^{\mathsf{T}} x \mathrm{d}t - 2 \int_{-\infty}^{\infty} \Phi(x)^{\mathsf{T}} \Phi(x) \mathrm{d}t \\ &= \int_{-\infty}^{\infty} \Big(-2ab x^{\mathsf{T}} x + (a+b) x^T \Phi(x) + (a+b) \Phi(x)^{\mathsf{T}} x - 2 \Phi(x)^{\mathsf{T}} \Phi(x) \Big) \mathrm{d}t \end{aligned}$$

which can be written as

$$\int_{-\infty}^{\infty} \begin{bmatrix} x(t) \\ \Phi(x(t)) \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} -2ab & a+b \\ a+b & -2 \end{bmatrix} \begin{bmatrix} x(t) \\ \Phi(x(t)) \end{bmatrix} \mathrm{d}t \geq 0$$

Applying Parseval's theorem [5]

$$= \frac{1}{2\pi} \Bigg[-2ab \int_{-\infty}^{\infty} \hat{x}^*(j\omega)\hat{x}(j\omega)\mathrm{d}\omega + (a+b) \int_{-\infty}^{\infty} \hat{x}^*(j\omega)\hat{\phi}(j\omega)\mathrm{d}\omega \\ +(a+b) \int_{-\infty}^{\infty} \hat{\phi}^*(j\omega)\hat{x}(j\omega)\mathrm{d}\omega - 2 \int_{-\infty}^{\infty} \hat{\phi}^*(j\omega)\hat{\phi}(j\omega)\mathrm{d}\omega \Bigg]$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Big(-2ab\hat{x}^*(j\omega)\hat{x}(j\omega) + (a+b)\hat{x}^*(j\omega)\hat{\phi}(j\omega) + (a+b)\hat{\phi}^*(j\omega)\hat{x}(j\omega) - 2\hat{\phi}^*(j\omega)\hat{\phi}(j\omega) \Big) \mathrm{d}\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \begin{bmatrix} \hat{x}(j\omega) \\ \hat{\phi}(j\omega) \end{bmatrix}^{*} \begin{bmatrix} -2ab & a+b \\ a+b & -2 \end{bmatrix} \begin{bmatrix} \hat{x}(j\omega) \\ \hat{\phi}(j\omega) \end{bmatrix} \mathrm{d}\omega \geq 0, \quad \forall x \in \mathcal{L}_2^m$$

We can prove the same result in the discrete-time case. ■

2.6.2 Sector-Bounded Dynamical Systems

Lemma 2.3 (Equivalent forms of sector-bounded dynamic systems)**.** *If $\mathcal{H}$ is a Hilbert space and the inner product $\langle \cdot, \cdot \rangle$ is associated with the norm $\| \cdot \|$, and if $\Phi : \mathcal{H} \to \mathcal{H}$ and $\Phi(0) = 0$ where 0 is the zero vector in $\mathcal{H}$, then assuming $b \geq a$, following two statements are equivalent:*

1. $\langle \Phi x - ax, \, \Phi x - bx \rangle_T \leq 0, \quad \forall T \geq 0, \, x \in \mathcal{H}.$
2. $\|\Phi x - \frac{a+b}{2} x\|_T \leq \frac{b-a}{2} \|x\|_T, \quad \forall T \geq 0, \, x \in \mathcal{H}$
3. $\int_{-\infty}^{\infty} \begin{bmatrix} x_T(t) \\ (\Phi x)_T(t) \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} -2ab & a+b \\ a+b & -2 \end{bmatrix} \begin{bmatrix} x_T(t) \\ (\Phi x)_T(t) \end{bmatrix} \mathrm{d}t \geq 0, \quad \forall T \geq 0, \quad \forall x \in L_{2e}^n.$

Proof.

1 $\iff$ 2**:** Proof taken from [5] for the non-cumulative version

$$\begin{aligned} 0 &\geq \langle \Phi x - ax, \, \Phi x - bx \rangle_T \\ &= \langle \Phi x, \, \Phi x \rangle_T - 2\frac{a+b}{2} \langle \Phi x, \, x \rangle_T + \left(\frac{(a+b)^2}{4} - \frac{(b-a)^2}{4} \right) \langle x, \, x \rangle_T \\ &= \left\| \Phi x - \frac{a+b}{2} x \right\|_T^2 - \frac{(b-a)^2}{4} \|x\|_T^2 \end{aligned}$$

[5] Assuming that system Φ is bounded and therefore the integral is defined.

$2 \iff 3$ Start from (2.6), square it and for all $T \geq 0$ and $x \in \mathcal{H}$ write it as

$$\left\langle \Phi x - \frac{a+b}{2}x,\, \Phi x - \frac{a+b}{2}x \right\rangle_T - \frac{(b-a)^2}{4} \langle x,\, x\rangle_T \leq 0$$
$$\Rightarrow -2\langle \Phi x,\, \Phi x\rangle_T - 2ab\langle x,\, x\rangle_T + (a+b)\langle x,\, \Phi x\rangle_T + (a+b)\langle x,\, \Phi x\rangle_T \geq 0$$

If we assume that $\mathcal{H} = \mathcal{L}_e^n$ and $x(t)$ and $(\Phi x)(t)$ are maps from $\mathbb{R}_+$ to $\mathbb{R}^n$, we can expand these inner-products. we will have

$$-2\int_{-\infty}^{T} (\Phi x)^T(t)\Phi x(t)\mathrm{d}t - 2ab\int_{-\infty}^{T} x^T(t)x(t)\mathrm{d}t + (a+b)\int_{-\infty}^{T} x^T(t)\Phi x(t)\mathrm{d}t + (a+b)\int_{-\infty}^{T} x^T(t)\Phi x(t)\mathrm{d}t$$
$$= \int_{-\infty}^{T} \begin{bmatrix} x(t) \\ (\Phi x)(t) \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} -2ab & a+b \\ a+b & -2 \end{bmatrix} \begin{bmatrix} x(t) \\ (\Phi x)(t) \end{bmatrix} \mathrm{d}t$$

■

Chapter 3

Generalized Necessary and Sufficient Robust Boundedness Results for Feedback Systems

Classical conditions for ensuring the robust stability of a system in feedback with a nonlinearity include passivity, small gain, circle, and conicity theorems. We present a generalized and unified version of these results in an arbitrary semi-inner product space, which avoids many of the technicalities that arise when working in traditional extended spaces. Our general formulation clarifies when the sufficient conditions for robust stability are also necessary, and we show how to construct worst-case scenarios when the sufficient conditions fail to hold. Finally, we show how our general result can be specialized to recover a wide variety of existing results, and explain how properties such as boundedness, causality, linearity, and time-invariance emerge as a natural consequence.

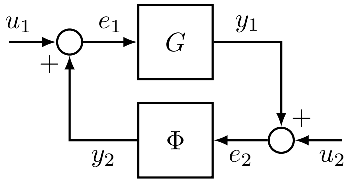
3.1 Introduction

Robust stability of interconnected systems has been a topic of interest for over 75 years, dating back to the seminal works of Lur'e [1], Zames [41,42], and Willems [6]. The standard input-output setup is illustrated in Fig. 3.1, where systems G and Φ are connected in feedback, and we seek conditions under which we can ensure the stability of the closed-loop map $(u_1, u_2) \to (y_1, y_2)$.

Robust stability is usually stated as a sufficient condition. For example: "Suppose $G \in S_1$ and $\Phi \in S_2$. If a certain condition holds on S_1 and S_2, then the interconnection of Fig. 3.1 is stable." In typical usage, a known system G is interconnected with some unknown, uncertain, or otherwise troublesome nonlinearity $\Phi \in S_2$, where S_2 is given. Then, ensuring stability of the interconnected system for any $\Phi \in S_2$ amounts to verifying the condition $G \in S_1$.

There are many robust stability results in the literature: passivity theory, the small-gain theorem, the circle criterion, graph separation, conic sector theorems, multiplier theory, dissipativity theory, and integral quadratic constraints.[1]

The aforementioned results provide *sufficient* conditions for robust stability. To reduce conservatism, one may ask conditions for robust stability that are *necessary and sufficient*. Such results typically take the



$$e_1 = u_1 + y_2 \tag{3.1a}$$

$$y_2 = \Phi e_2 \tag{3.1b}$$

$$e_2 = u_2 + y_1 \tag{3.1c}$$

$$y_1 = G e_1 \tag{3.1d}$$

Figure 3.1: Feedback interconnection of systems G and Φ.

form: "Suppose S_1 and S_2 satisfy a certain condition. Then, $G \in S_1$ if and only if the interconnection of Fig. 3.1 is stable for all $\Phi \in S_2$." Again, there are numerous examples in the literature, such as necessary and sufficient versions of the passivity, small-gain, and circle theorems.[1]

The reason for the wide variety of robust stability results is that different assumptions can be made about G and Φ, and the sets S_1 and S_2 can be defined in many different ways. A natural question to ask, which forms the basis of our present work, is whether the multitude of existing results can be viewed as consequences of a single more general result. We answer in the affirmative.

Main contributions: In Section 3.2, we present a robust boundedness result involving interconnected *relations* over a general *semi-inner product space*, where there need not exist a notion of time. Our result (Theorem 3.1) does not assume linearity or even boundedness of G or Φ, and avoids the technicalities typically associated with *causality*, *stability*, and *well-posedness*. Under mild conditions, our sufficient condition for robust boundedness is also necessary, and we provide a constructive proof.

In Section 3.3, we specialize our result to standard extended spaces of time-domain signals (e.g., $L_{2\mathrm{e}}$ or $\ell_{2\mathrm{e}}$), which reveals the way in which the aforementioned technicalities arise. We also explain why necessity is more difficult to achieve in this setting, and why stronger assumptions (e.g., linearity and time-invariance of G) are often required.

A key observation, explained in Section 3.4, is that *sufficient-only* results are often[2] a direct consequence of a corresponding necessary-and-sufficient counterpart, meaning that there is nothing to be gained by stating a sufficient-only version.

3.1.1 Related Work

In Table 3.1, we provide a summary of existing robust stability results. In the "Implication" column, we distinguish between *sufficient-only* results ($\Longrightarrow$) and *necessary-and-sufficient* results ($\Longleftrightarrow$).

[1] Detailed references can be found in Section 3.1.1 and Table 3.1.

[2] We make this claim for results involving conicity constraints. These results include: passivity, small-gain, circle criterion, conicity, and extended conicity. See Section 3.1.1 and Table 3.1 for details.

Table 3.1: Literature Review of robust stability results involving an interconnection of two systems (refer to Fig. 3.1). The first group of rows are sufficient-only results (Implication: $\Longrightarrow$). The second group of rows are necessary and sufficient (Implication: $\Longleftrightarrow$). For constraints on $G \in \mathcal{C}_G$ and $\Phi \in \mathcal{C}_\Phi$ (refer to Fig. 3.1) we denote linear (L), nonlinear (NL), time-varying (TV), time-invariant (TI), static (S), and fading-memory (FM). For example, "NLTV" indicates nonlinear and time-varying. For vector spaces, the symbols $L_{2\mathrm{e}}$, $\ell_{2\mathrm{e}}$, $\mathcal{L}_{2\mathrm{e}}$ denote extended spaces (see Section 2.1.1) and s.i.p.s. denotes a semi-inner product space. The final column indicates whether the proof of the converse direction ($\Longleftarrow$), if applicable, explicitly constructs a worst-case Φ when the conditions on G are violated.

Reference	**Result Type**	$\mathcal{C}_G$	$\mathcal{C}_\Phi$	**Implication**	**Vector Space**	**Constructive?**
Vidyasagar [3, §6.6.(1,58)]	passivity & small gain	NLTV	NLTV	$\Longrightarrow$	$\mathcal{L}_{2\mathrm{e}}$	
Zames [41, Thm. 1–3]	conic	NLTV	NLTV	$\Longrightarrow$	$\mathcal{L}_{2\mathrm{e}}$	
Bridgeman & Forbes [12]	conic	NLTV	NLTV	$\Longrightarrow$	$\mathcal{L}_{2\mathrm{e}}$	
Zames [42, §3–4]	circle & multipliers	LTI	NLTVS	$\Longrightarrow$	$L_{2\mathrm{e}}$	
Desoer & Vidyasagar [5]	multipliers	NLTV	NLTV	$\Longrightarrow$	$\mathcal{L}_{2\mathrm{e}}$	
Teel et al. [43]	graph separation	NLTV	NLTV	$\Longrightarrow$	$\mathcal{L}_{2\mathrm{e}}$	
Willems [6]	dissipativity	NLTV	NLTV	$\Longrightarrow$	$\mathcal{L}_{2\mathrm{e}}$	
Pfifer & Seiler [44]	dissipativity	LTI	NLTV	$\Longrightarrow$	$L_{2\mathrm{e}}$	
Megretski & Rantzer [13]	IQC	LTI	NLTVS	$\Longrightarrow$ [†]	$L_{2\mathrm{e}}$	
Vidyasagar [3, §6.6.(112,126)]	small gain & circle	LTI	NLTV	$\Longleftrightarrow$	$L_{2\mathrm{e}}$	Yes
Khong & van der Schaft [45, Thm. 3]	passivity & small gain	LTI	LTV	$\Longleftrightarrow$	$L_{2\mathrm{e}}$	No
Zhou et al. [46, Thm. 9.1]	small gain	LTI	LTI	$\Longleftrightarrow$	$L_{2\mathrm{e}}$	Yes
Khong & Kao [47, Thm. 1]	IQC	LTI	LTI	$\Longleftrightarrow$	$L_{2\mathrm{e}}$	Yes
Shamma [48, Thm. 3.2]	small gain	NLTVFM	NLTVFM	$\Longleftrightarrow$	$\ell_{2\mathrm{e}}$	Yes
Cyrus & Lessard [49]	conic	L	NL	$\Longleftrightarrow$	s.i.p.s.	No
Present work	conic	NL	NL	$\Longleftrightarrow$	s.i.p.s.	Yes

[†] The authors in [13] mention that their sufficient condition for robust stability is also necessary in the sense that a result in spirit of Lemma 3.1 holds via a suitable application of the S-lemma [50].

Sufficient results Classical sufficient results include the passivity, small-gain, and circle theorems[3]. These results are mutually related via a loop-shifting transformation [15], and were generalized to *conic sectors* [41, 42, 51].

Beyond conic sector constraints, graph separation [43, 52] allows for nonlinear constraints, while multiplier theory [5], dissipativity [6], and integral quadratic constraints (IQCs) [13, 44, 53] allow for dynamic or time-varying constraints. There have also been several works discussing how these various frameworks are related [54–56].

Necessary and sufficient results The classical passivity, small-gain, and circle theorems are only sufficient when Φ is assumed to be memoryless (but still possibly time-varying) [10, 57]. However, necessity holds if Φ is allowed to have memory, e.g. when Φ is a dynamical system [3, Thm. 6.6.126].

The majority of necessary and sufficient robust stability results assume that G is linear. For example, the passivity and small gain results of Vidyasagar [3, §6.6(112,126)] and Khong et al. [45, Thm. 3] assume G is linear and time-invariant (LTI). Meanwhile, the small gain result of Zhou et al. [46, Thm. 9.1] and the recent converse IQC result of Khong et al. [47] make the stronger assumption that *both* G and Φ are LTI. Finally, Shamma's small gain theorem [48, Thm. 3.2] holds when both G and Φ are nonlinear and time-varying, but requires an additional assumption of *fading memory*, which effectively allows the system response to be approximated by that of a linear system.

One reason for requiring linearity of G in necessary-and-sufficient results is that proving the necessary direction is often done via the S-lemma [50, 58], which requires linearity. We go into more detail on this point in Section 3.2.2. It turns out that for passivity, small gain, circle, and conicity results, linearity is not actually required, even though it is often assumed. Our main result (Theorem 3.1) provides a unified robust stability result without any linearity requirements.

An earlier version of Theorem 3.1 [49] used the S-lemma and therefore assumed linearity and was nonconstructive.

3.1.2 Notation

Semi-inner products A *semi-inner product space* is a vector space $\mathcal{V}$ over a field $\mathbb{F}$ equipped with a semi-inner product[4] $\langle \cdot,\, \cdot \rangle$, which is an inner product whose associated norm is a seminorm. In other words, $\|x\| := \sqrt{\langle x,\, x\rangle} \geq 0$ for all $x \in \mathcal{V}$, but $\|x\| = 0$ need not imply that $x = 0$.

[3] The circle criterion in the literature typically refers to the case when G in Figure 3.1 is assumed to be an LTI system.

[4] We use the convention that a semi-inner product is linear in its second argument, so $\langle x,\, ay + bz\rangle = a\,\langle x,\, y\rangle + b\,\langle x,\, z\rangle$ for all $x, y, z \in \mathcal{V}$ and $a, b \in \mathbb{F}$. Also, $\langle x,\, y\rangle = \overline{\langle y,\, x\rangle}$.

Relations A *relation* R on $\mathcal{V}$ is a subset of the product space $R \subseteq \mathcal{V} \times \mathcal{V}$. We write $\mathscr{R}(\mathcal{V})$ to denote the set of all relations on $\mathcal{V}$. The ***domain*** of R is defined as $\operatorname{dom}(R) := \{x \in \mathcal{V} \mid (x, y) \in R \text{ for some } y \in \mathcal{V}\}$. For any $x \in \operatorname{dom}(R)$, we write Rx to denote any $y \in \mathcal{V}$ such that $(x, y) \in R$.

We define $\mathcal{V}^2$ to be the augmented vectors $u =: \left(\begin{smallmatrix} u_1 \\ u_2 \end{smallmatrix}\right)$ where $u_1, u_2 \in \mathcal{V}$. We overload matrix multiplication in $\mathcal{V}^2$, specifically, for any $\xi, \zeta \in \mathcal{V}^2$ and any matrix $N \in \mathbb{F}^{2\times 2}$,

$$N\xi = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} := \begin{bmatrix} N_{11}\xi_1 + N_{12}\xi_2 \\ N_{21}\xi_1 + N_{22}\xi_2 \end{bmatrix} \in \mathcal{V}^2.$$

Likewise, inner products in $\mathcal{V}^2$ have the interpretation

$$\langle \xi, \zeta \rangle = \left\langle \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \right\rangle := \langle \xi_1, \zeta_1 \rangle + \langle \xi_2, \zeta_2 \rangle .$$

The closed-loop system of Fig. 3.1 defines the following relations, which characterize pairs of consistent signals.

$$\begin{aligned} R_{uy} &:= \left\{ (u, y) \in \mathcal{V}^2 \times \mathcal{V}^2 \;\middle|\; (3.1) \text{ holds for some } e \in \mathcal{V}^2 \right\}, \\ R_{ue} &:= \left\{ (u, e) \in \mathcal{V}^2 \times \mathcal{V}^2 \;\middle|\; (3.1) \text{ holds for some } y \in \mathcal{V}^2 \right\}. \end{aligned}$$

3.2 Results for Semi-Inner Product Spaces

The main result of this section is a robust boundedness theorem defined over a general semi-inner product space. We consider the setup of Fig. 3.1, where $G \in \mathcal{C}_G$ and $\Phi \in \mathcal{C}_\Phi$ belong to arbitrary sets of (possibly nonlinear) relations.

Theorem 3.1. *Let $\mathcal{V}$ be a semi-inner product space and let $M = M^* \in \mathbb{F}^{2\times 2}$ be indefinite. Suppose $G \in \mathcal{C}_G$ and consider the three following statements.*

1. *There exists $N = N^* \in \mathbb{F}^{2\times 2}$ satisfying $M + N \prec 0$ such that G satisfies*

$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle \geq 0 \quad \textit{for all } \xi \in \operatorname{dom}(G). \tag{3.3}$$

2. *There exists $\gamma > 0$ such that for all (u, y, e), if*

$$\left\langle \begin{bmatrix} e_2 \\ y_2 \end{bmatrix}, M \begin{bmatrix} e_2 \\ y_2 \end{bmatrix} \right\rangle \geq 0 \tag{3.4}$$

and (3.1a), (3.1c), (3.1d) *are satisfied, then $\|y\| \leq \gamma \|u\|$.*

3. *There exists $\gamma > 0$ such that for all $\Phi \in \mathcal{C}_\Phi$, if*

$$\left\langle \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix}, M \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix} \right\rangle \geq 0 \quad \textit{for all } \xi \in \operatorname{dom}(\Phi), \tag{3.5}$$

then for all $(u, y) \in R_{uy}$, the following bound holds

$$\|y\| \leq \gamma \|u\|. \tag{3.6}$$

The following equivalences hold: (1) $\iff$ (2) $\implies$ (3).

In Theorem 3.1, Item (1) is a property of subsystem G, while Item (3) is a statement about the robust boundedness of the closed loop when G is interconnected with any $\Phi \in \mathcal{C}_\Phi$. The intermediate statement (2) is similar to (3), but it concerns robustness with respect to the signals (u, y, e) rather than Φ.

We can view Theorem 3.1 as a *sufficient* condition for robust boundedness because it proves that (1) $\implies$ (3). Since Theorem 3.1 holds for arbitrary choices of $\mathcal{C}_G$ and $\mathcal{C}_\Phi$, it generalizes the results presented in the first half of Table 3.1. We discuss the details of how to specialize Theorem 3.1 in Section 3.3.

We make several remarks about Theorem 3.1.

Remark 3.1. *Equation* (3.6) *can be stated in terms of* (u, e) *instead of* (u, y)*. Specifically, it is easy to show that* (3.6) *holds for all* $(u, y) \in R_{uy}$ *if and only if there exists some* $\bar{\gamma} > 0$ *such that* $\|e\| \leq \bar{\gamma}\|u\|$ *holds for all* $(u, e) \in R_{ue}$*.*

Remark 3.2. *In Theorem 3.1,* M *is assumed to be indefinite. If* M *is semidefinite instead, it will generally lead to results that are either trivial or vacuous statements, that is, Item (1) and Item (3) are either always true or always false.*

Remark 3.3. *In Item (1), we can equivalently replace* N *by* $-M - \varepsilon I$ *and modify the statement preceding* (A.1) *to: "There exists some* $\varepsilon > 0$ *such that* G *satisfies* (A.1)*". We chose the form with* M *and* N *to make Theorem 3.1 more symmetric.*

Remark 3.4. *Theorem 3.1 can be further generalized to the case where* $G \in \mathcal{C}_G \subseteq \mathcal{R}(\mathcal{V}^n, \mathcal{V}^m)$ *(the set of relations on* $\mathcal{V}^n \times \mathcal{V}^m$*) and* $\Phi \in \mathcal{C}_\Phi \subseteq \mathcal{R}(\mathcal{V}^m, \mathcal{V}^n)$*. In this case,* $M, N \in \mathbb{F}^{(m+n)\times(m+n)}$ *would be block* 2×2 *matrices.*

Remark 3.5. *In Theorem 3.1, both* G *and* Φ *are relations rather than operators. All relations are invertible, so the closed-loop relations* R_{uy} *and* R_{ue} *are always well-defined, but may be* empty*. In such a case,* (3.6) *is vacuously true. One way to ensure that Theorem 3.1 is not vacuous is to require the* well-posedness *assumption that* R_{uy} *(equivalently* R_{ue}*) is non-empty.*

Since Theorem 3.1 is expressed in a general semi-inner product space, there need not exist a notion of time and concepts such as causality and stability need not apply. We will see in Section 3.3 how concepts such as causality, stability, and well-posedness emerge when Theorem 3.1 is specialized to extended spaces of time-domain signals (Lebesgue spaces) and G and Φ are *operators* rather than relations.

3.2.1 Proof of Sufficiency for Theorem 3.1

We begin by showing that (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) in Theorem 3.1 for any choice of $\mathcal{C}_G$ and $\mathcal{C}_\Phi$. This proof is similar to [59, Thm. 1]. Pick any (u, y, e) such that (3.1a), (3.1c), (3.1d), and (A.2) are satisfied. let $\xi = e_1$ in (A.1). Using (3.1) to eliminate e_1, e_2, Equations (A.1) and (A.2) become:

$$\left\langle \begin{bmatrix} y_1 \\ u_1 + y_2 \end{bmatrix}, N \begin{bmatrix} y_1 \\ u_1 + y_2 \end{bmatrix} \right\rangle \geq 0 \quad \text{and} \quad \left\langle \begin{bmatrix} u_2 + y_1 \\ y_2 \end{bmatrix}, M \begin{bmatrix} u_2 + y_1 \\ y_2 \end{bmatrix} \right\rangle \geq 0.$$

Sum the two inequalities above and collect terms to obtain

$$\left\langle \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, (M+N) \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\rangle + 2 \left\langle \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \begin{bmatrix} N_{12} & M_{11} \\ N_{22} & M_{21} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right\rangle + \left\langle \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} N_{22} & 0 \\ 0 & M_{11} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right\rangle \geq 0.$$

Since $M + N \prec 0$ by assumption, There exists $\eta > 0$ such that $M + N \preceq -\eta I$. Applying this inequality together with Cauchy–Schwarz[5], we obtain

$$-\eta\|y\|^2 + 2r\|y\|\|u\| + q\|u\|^2 \geq 0, \tag{3.7}$$

where $r := \left\| \left[\begin{smallmatrix} N_{12} & M_{11} \\ N_{22} & M_{21} \end{smallmatrix}\right] \right\|$ and $q := \left\| \left[\begin{smallmatrix} N_{22} & 0 \\ 0 & M_{11} \end{smallmatrix}\right] \right\|$ are standard spectral norms. Dividing by η and completing the square in (3.7), we obtain $\left(\|y\| - \frac{r}{\eta}\|u\|\right)^2 \leq \frac{r^2+\eta q}{\eta^2}\|u\|^2$, which can be rearranged to establish (2) with $\gamma = \frac{1}{\eta}\left(r + \sqrt{r^2 + \eta q}\right)$.

To prove (3), consider some $\Phi \in \mathcal{C}_\Phi$ for which (3.5) holds. Next, pick $(u, y) \in R_{uy}$ so that there exists (u, y, e) satisfying (3.1). In particular, (3.1b) holds, so setting $\xi = e_2$ in (3.5), we obtain (A.2) and the rest of the proof is the same as above. ■

3.2.2 Proof of Partial Necessity for Theorem 3.1

A popular approach for proving (1) $\Longleftarrow$ (2) is to use a lossless S-lemma as in [45, Thm. 3] and [13]. However, the S-lemma [50, 58] comes with a drawback: the set of signals (u, y, e) that satisfy the loop equations (3.1a), (3.1c), (3.1d) must be a *subspace*, which requires for example that G be linear. If we assume G is linear, we can prove (1) $\Longleftarrow$ (2) by adapting the S-lemma for inner product spaces due to Hestenes [61, Thm. 7.1, p. 354] and using a technique similar to that used in [45]. Details of this approach may be found in [49] and Appendix A.

It turns out the linearity assumption on G can be dropped entirely if we adopt a different proof approach. To this effect, we will prove the contrapositive $\neg(1) \Longrightarrow \neg(2)$ by directly constructing signals (y, u, e) that violate the boundedness condition when (1) fails to hold. Unlike the S-lemma, this approach does not require linearity of G and has the benefit of being constructive, so it produces *worst-case* signals (u, y, e). We state the result in the following lemma.

Lemma 3.1 (worst-case signals)**.** *Consider the setting of Theorem 3.1. Suppose that for any N satisfying $M + N \prec 0$, there exists $\xi \in \operatorname{dom}(G)$ such that*

$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle < 0.$$

Then, for all $\gamma > 0$, there exists (u, y, e) such that:

1. *Equations* (3.1a)*,* (3.1c)*, and* (3.1d) *hold.*
2. $\left\langle \begin{bmatrix} e_2 \\ y_2 \end{bmatrix}, M \begin{bmatrix} e_2 \\ y_2 \end{bmatrix} \right\rangle \geq 0.$

[5] A proof of the Cauchy–Schwarz inequality for general semi-inner product spaces may be found in [60, §1.4].

3. $\|y\| > \gamma\|u\|$.

Proof. The proof is constructive and explicitly produces the signals (u, y, e) as functions of ξ and γ. See Section 3.6.1.1 for a detailed proof. ■

Although Theorem 3.1 is only a sufficient result, the intermediate Item (2) in Theorem 3.1 provides an important clue toward finding necessary-and-sufficient results. Specifically, we need only focus on proving (2) $\Longleftarrow$ (3) and we can ignore (1) entirely. This is the topic of the next section.

3.2.3 Toward Achieving Necessity

Theorem 3.1 states that for arbitrary choices of $\mathcal{C}_G$ and $\mathcal{C}_\Phi$, the items satisfy the partial equivalence (1) $\iff$ (2) $\implies$ (3). In general, the three items will not be equivalent. However, there are special choices of $\mathcal{C}_G$ and $\mathcal{C}_\Phi$ that ensure the missing implication (2) $\Longleftarrow$ (3) holds and therefore the three items in Theorem 3.1 become equivalent.

Definition 3.1. *We say that a pair of constraint sets* $(\mathcal{C}_G, \mathcal{C}_\Phi)$ achieves necessity *if such a choice implies that (2)* $\Longleftarrow$ *(3) in Theorem 3.1.*

Our first observation is that shrinking $\mathcal{C}_G$ or enlarging $\mathcal{C}_\Phi$ preserves the validity of the necessity direction. This leads us to the following proposition.

Proposition 3.1. *If* $(\mathcal{C}_G, \mathcal{C}_\Phi)$ *achieves necessity, then* $(\mathcal{C}'_G, \mathcal{C}'_\Phi)$ *also achieves necessity for any* $\mathcal{C}'_G \subseteq \mathcal{C}_G$ *and* $\mathcal{C}'_\Phi \supseteq \mathcal{C}_\Phi$.

Proof. If Item (3) of Theorem 3.1 holds for $\mathcal{C}'_\Phi$, then it must also hold for $\mathcal{C}_\Phi \subseteq \mathcal{C}'_\Phi$ since the condition ranges over a smaller set of candidate Φ's. Therefore, if $(\mathcal{C}_G, \mathcal{C}_\Phi)$ achieves necessity, then so does $(\mathcal{C}_G, \mathcal{C}'_\Phi)$. Theorem 3.1 is a statement about $G \in \mathcal{C}_G$, so if $G \in \mathcal{C}'_G \subseteq \mathcal{C}_G$, then Theorem 3.1 clearly still holds. Thus, if $(\mathcal{C}_G, \mathcal{C}'_\Phi)$ achieves necessity, then so does $(\mathcal{C}'_G, \mathcal{C}'_\Phi)$. This completes the proof. ■

In light of Proposition 3.1, our goal should be to find the *smallest* $\mathcal{C}_\Phi$ and *largest* $\mathcal{C}_G$ such that $(\mathcal{C}_G, \mathcal{C}_\Phi)$ achieves necessity. In the remainder of this section, we present two different ways of achieving necessity through specific choices of $\mathcal{C}_G$ and $\mathcal{C}_\Phi$.

Our first result is that Theorem 3.1 holds if the constraint on $\mathcal{C}_\Phi$ is removed entirely.

Theorem 3.2 (unconstrained case)**.** *The constraint set* $(\mathcal{C}_G, \mathcal{C}_\Phi)$ *achieves necessity when* $\mathcal{C}_G = \mathcal{C}_\Phi = \mathscr{R}(\mathcal{V})$.

Proof. We proceed by contradiction. Suppose (2) fails. Then for all $\gamma > 0$, there exists (u, y, e) satisfying (A.2) and (3.1a), (3.1c), (3.1d) such that $\|y\| > \gamma\|u\|$. Since $\mathcal{C}_\Phi = \mathscr{R}(\mathcal{V})$, we can pick $\Phi = \{(e_2, y_2)\}$, a singleton relation. Then both (3.1b) and (3.5) hold trivially and so (3) fails, as required. ■

Theorem 3.2 may not be particularly satisfying because it requires defining Φ as a singleton relation. A more common use case is when Φ must be defined for all $\mathrm{dom}(\Phi) = \mathcal{V}$.[6] Our second result states that necessity can be achieved by *linear* relations, which we now define.

Definition 3.2 (linear relation)**.** *Let $\mathcal{V}$ be a semi-inner product space over a field $\mathbb{F}$. Suppose $x_1, x_2, y_1, y_2 \in \mathcal{V}$ and $\alpha_1, \alpha_2 \in \mathbb{F}$. A relation $R \in \mathscr{R}(\mathcal{V})$ is* linear *if for all $(x_1, y_1) \in R$ and $(x_2, y_2) \in R$, we have $(\alpha_1 x_1 + \alpha_2 x_2, \alpha_1 y_1 + \alpha_2 y_2) \in R$. We let $\mathscr{L}(\mathcal{V}) \subseteq \mathscr{R}(\mathcal{V})$ denote the set of all linear relations.*

Theorem 3.3 (linear case)**.** *The constraint set $(\mathcal{C}_G, \mathcal{C}_\Phi)$ achieves necessity when $\mathcal{C}_G = \mathscr{R}(\mathcal{V})$ and $\mathcal{C}_\Phi \supseteq \mathscr{L}(\mathcal{V})$.*

Proof. The proof of Theorem 3.3 is constructive and explicitly produces a worst-case $\Phi \in \mathscr{L}(\mathcal{V})$. See Section 3.6.1.2 for a detailed proof. ■

Theorem 3.2 and 3.3 both provide conditions that ensure necessity of Theorem 3.1. In both cases, $\mathcal{C}_G = \mathscr{R}(\mathcal{V})$, so there are no constraints on G; it could be nonlinear, for example.

3.3 Specialization to Extended Spaces

In this section, we specialize Theorem 3.1 to the popular use case where the loop signals (u, y, e) in Fig. 3.1 are functions of time (either discrete or continuous) and the systems G and Φ are *causal operators* rather than relations. This specialization introduces the familiar concepts of *causality*, *well-posedness*, and *input-output stability.*

This section mirrors Section 3.2; after some notational preliminaries, we present a corollary to Theorem 3.1 that holds for causal operators and describe conditions that ensure necessity.

In this chapter,

3.3.1 Main Results for Extended Spaces

Before specializing Theorem 3.1 to the Lebesgue spaces defined in Section 2.1.1, we must discuss how well-posedness and causality fit into the picture.

3.3.1.1 Well-posedness

Assuming G and Φ are relations, as we do in Theorem 3.1, is not unprecedented in the literature [3, 41, 45, 52, 62, 63]. As mentioned in Remark 3.5, this strategy ensures well-posedness of any interconnection. However, the closed-loop relations R_{uy} (equivalently R_{ue}) may be empty. When G and Φ are assumed to be operators instead of relations, then well-posedness must either be assumed or proved. Specifically, we need an assurance of the existence and uniqueness of solutions e and y for all choices of u.

[6] Relations $\Phi \in \mathscr{R}(\mathcal{V})$ that satisfy $\mathrm{dom}(\Phi) = \mathcal{V}$ are known as *serial* or *left-total*. They are also called *multi-valued functions.*

3.3.1.2 Causality

When working in extended spaces such as $\mathcal{L}_{2\mathrm{e}}$, a common assumption is that G and Φ are causal operators [3, 5, 13, 42, 46]. In specializing Theorem 3.1 to extended spaces, we will let $\mathcal{C}_G$ and $\mathcal{C}_\Phi$ be sets of causal operators on $\mathcal{L}_{2\mathrm{e}}$. Then, since a well-posed interconnection of causal maps is causal [62, Prop. 1.2.14], the closed-loop map will be causal.

Although Theorem 3.1 can in principle be specialized to $\mathcal{V} = \mathcal{L}_2$, this does not yield a fruitful result. In particular, Item (3) of Theorem 3.1 would state that for all Φ satisfying the appropriate constraints and for which R_{uy} is non-empty, we would have $\|y\| \leq \gamma\|u\|$. Since the typical use case of Theorem 3.1 is to prove that the closed-loop map $u \mapsto y$ is bounded, this shifts all the burden onto proving R_{uy} is non-empty (well-posedness). In other words, assuming well-posedness amounts to assuming that which we seek to prove.

To resolve the aforementioned problem, we instead specialize Theorem 3.1 to $\mathcal{V} = \mathcal{L}_{2\mathrm{e}}$ and let $\mathcal{C}_G \subseteq \mathscr{F}(\mathcal{L}_{2\mathrm{e}})$ and $\mathcal{C}_\Phi \subseteq \mathscr{F}(\mathcal{L}_{2\mathrm{e}})$ be arbitrary constraint sets. This leads to the following main result of this section.

Corollary 3.1 (robust input-output stability on $\mathcal{L}_{2\mathrm{e}}$)**.** *Let $M = M^* \in \mathbb{F}^{2\times 2}$ be indefinite. Suppose $G \in \mathcal{C}_G$ and consider the three following statements.*

1. *There exists $N = N^* \in \mathbb{F}^{2\times 2}$ satisfying $M + N \prec 0$ such that for all $\xi \in \mathcal{L}_{2e}$ and $T \geq 0$, G satisfies*
$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle_T \geq 0. \tag{3.8}$$

2. *There exists $\gamma > 0$ such that for all (u, y, e), if*
$$\left\langle \begin{bmatrix} e_2 \\ y_2 \end{bmatrix}, M \begin{bmatrix} e_2 \\ y_2 \end{bmatrix} \right\rangle_T \geq 0 \quad \text{for all } T \geq 0 \tag{3.9}$$
and (3.1a)*,* (3.1c)*,* (3.1d) *hold and $u \in \mathcal{L}_2$, then $\|y\| \leq \gamma\|u\|$.*

3. *There exists $\gamma > 0$ such that for all $\Phi \in \mathcal{C}_\Phi$ where the interconnection of Fig. 3.1 is well-posed, if for all $T \geq 0$ and $\xi \in \mathcal{L}_{2e}$ we have*
$$\left\langle \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix}, M \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix} \right\rangle_T \geq 0, \tag{3.10}$$
then for all $y = R_{uy}u$ with $u \in \mathcal{L}_2$, we have
$$\|y\| \leq \gamma\|u\|. \tag{3.11}$$

The following equivalences hold: (1) $\iff$ (2) $\implies$ (3).

Proof. See Section 3.6.2.1. ■

Corollary 3.1 is similar to Theorem 3.1 in that no assumptions are made on $\mathcal{C}_G$ and $\mathcal{C}_\Phi$. The operators G and Φ may be nonlinear or even unbounded. The requirements (3.8) and (3.10) involve the truncated T norms, so they are defined even for unbounded signals.[7] Critically, the conclusion (3.11) states that when $u \in \mathcal{L}_2$, we have $y \in \mathcal{L}_2$, which is a statement about *input-output stability.*

Corollary 3.1 is a specialization of Theorem 3.1 to $\mathcal{V} = \mathcal{L}_{2\mathrm{e}}$ and therefore it is in the spirit of other time-domain results such as the conic sector theorem [41], and extended conic sector theorem [51].

As with general semi-inner product spaces, we can inquire about conditions on $\mathcal{C}_G$ and $\mathcal{C}_\Phi$ for which (2) $\Longleftarrow$ (3) and the result becomes necessary and sufficient for input-output stability. This is shown in Theorem 3.4.

Definition 3.3. *We say that a pair of constraint sets* $(\mathcal{C}_G, \mathcal{C}_\Phi)$ achieves necessity on $\mathcal{L}_{2\mathrm{e}}$ *if such a choice implies that (2)* $\Longleftarrow$ *(3) in Corollary 3.1.*

As in the semi-inner product setting, shrinking $\mathcal{C}_G$ or enlarging $\mathcal{C}_\Phi$ preserves the validity of the necessity direction. So Proposition 3.1 also holds on $\mathcal{L}_{2\mathrm{e}}$ when $\mathcal{C}_G$ and $\mathcal{C}_\Phi$ are subsets of $\mathscr{F}(\mathcal{L}_{2\mathrm{e}})$, the causal operators on $\mathcal{L}_{2\mathrm{e}}$.

Finding constraint sets that achieve necessity on $\mathcal{L}_{2\mathrm{e}}$ is more challenging than in the general semi-inner product setting because we require Φ to be a causal operator. Indeed, Theorem 3.2 constructs a worst-case Φ using a singleton relation, which is not a valid operator. Likewise, Theorem 3.3 constructs a linear worst-case Φ from the worst-case signals (e_2, y_2), but the resulting Φ is generally not causal.

One way to eliminate these difficulties is to make further assumptions on G. Specifically, we will assume G is LTI, which allows for the following equivalence between inner products and frequency domain inequalities (FDIs).

Lemma 3.2. *Let* $N = N^* \in \mathbb{F}^{2\times 2}$ *be indefinite and let* $P \in \mathbb{F}^{2\times 2}$ *diagonalize* N*, with* $N = P^* \left[\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right] P$*. Suppose* $G \in \mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2e})$ *with frequency response* $\hat{G}$*. The following two statements are equivalent.*

1. *For all* $T \geq 0$ *and for all* $\xi \in \mathcal{L}_{2e}$*, we have*

$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle_T \geq 0. \tag{3.12}$$

2. *The following frequency-domain inequality (FDI) holds:*

$$\begin{bmatrix} \hat{G}(\omega) \\ 1 \end{bmatrix}^* N \begin{bmatrix} \hat{G}(\omega) \\ 1 \end{bmatrix} \geq 0 \quad \text{for almost all } \omega \in \mathbb{R} \tag{3.13}$$

and $(P_{11}G + P_{12})(P_{21}G + P_{22})^{-1}$ *maps* $\mathcal{L}_2 \to \mathcal{L}_2$*.*

[7] Signals in $\mathcal{L}_{2\mathrm{e}}$ may be unbounded, but they cannot have finite escape time.

Proof. See Section 3.6.2.2 for a detailed proof. ■

We can now prove a version of Theorem 3.3 when G is constrained to be LTI.

Theorem 3.4 (LTI case)**.** *The constraint set* $(\mathcal{C}_G, \mathcal{C}_\Phi)$ *achieves necessity on* $\mathcal{L}_{2e}$ *when* $\mathcal{C}_G \subseteq \mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2e}) \subseteq \mathcal{C}_\Phi$.

Proof. The proof of Theorem 3.4 is constructive and inspired by Vidyasagar's necessary and sufficient circle criterion [3, Thm. 6.6.126]. See Section 3.6.2.3 for the proof. ■

Remark 3.6. *As noted in Section 3.2.2, if we assume that G is linear, the S-lemma can be used to prove the implication (1) ⟸ (2) in Corollary 3.1. However, this implication still holds even when G is nonlinear. Theorem 3.4 also assumes that G is LTI (in the* $\mathcal{L}_{2e}$ *setting), but instead proves the implication (2) ⟸ (3), for which the S-lemma cannot be used. Linearity is only used the* $\mathcal{L}_{2e}$ *setting to deal with the requirement that* Φ *be causal. In the semi-inner product setting, linearity of G is not required (see Theorems 3.2 and 3.3).*

We can combine Theorem 3.4 with Corollary 3.1 and Lemma 3.2 to obtain a necessary and sufficient condition relating frequency-domain properties of G with robust closed-loop stability. We diagonalize M instead of N for convenience, but the result can be formulated equivalently either way.

Theorem 3.5 (LTI case in the frequency domain)**.** *Let* $M = M^* \in \mathbb{F}^{2\times 2}$ *be indefinite and let* $P \in \mathbb{F}^{2\times 2}$ *diagonalize* M *with* $M = P^* \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right] P$. *Suppose* $G \in \mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2e})$ *has frequency response* $\hat{G}$ *and let* $\mathcal{C}_\Phi \supseteq \mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2e})$. *The following two statements are equivalent.*

1. *There exists* $N = N^* \in \mathbb{F}^{2\times 2}$ *satisfying* $M + N \prec 0$ *such that the following FDI holds for almost all* $\omega \in \mathbb{R}$*:*
$$\begin{bmatrix} \hat{G}(\omega) \\ 1 \end{bmatrix}^* N \begin{bmatrix} \hat{G}(\omega) \\ 1 \end{bmatrix} \geq 0, \tag{3.14}$$
and $(P_{11}G + P_{12})(P_{21}G + P_{22})^{-1}$ *maps* $\mathcal{L}_2 \to \mathcal{L}_2$.

2. *There exists* $\gamma > 0$ *such that for all* $\Phi \in \mathcal{C}_\Phi$ *where the interconnection of Fig. 3.1 is well-posed, if for all* $T \geq 0$ *and* $\xi \in \mathcal{L}_{2e}$ *we have*
$$\left\langle \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix}, M \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix} \right\rangle_T \geq 0, \tag{3.15}$$
then for all $y = R_{uy}u$ *with* $u \in \mathcal{L}_2$, *we have*
$$\|y\| \leq \gamma \|u\|. \tag{3.16}$$

When M and N are suitably chosen in Theorem 3.5, the frequency domain condition (3.14) can take on a familiar form, such as the classical circle criterion. In Section 3.4, we describe how different choices of M and N can be used to recover existing results in the literature.

The condition that $(P_{11}G+P_{12})(P_{21}G+P_{22})^{-1}$ maps $\mathcal{L}_2 \to \mathcal{L}_2$ is equivalent to $(P_{11}\mathbf{G}+P_{12})(P_{21}\mathbf{G}+P_{22})^{-1}$ being stable, where $\mathbf{G}$ is the transfer function of G. If we only have access to the frequency response $\hat{G}$, then the condition can be verified using the Nyquist criterion instead. Several existing results in the literature are expressed in this way, see for example [5, Thm. V.2.10], [36, Thm. 7.2], and [3, Thm. 6.6.126].

The frequency-domain condition (3.14) can be verified graphically, or if we have a state space representation for G, we can use the Kalman–Yakubovich–Popov (KYP) lemma to transform (3.14) into an equivalent linear matrix inequality (LMI), which admits a numerically tractable solution. We state the KYP lemma here (in the general MIMO case) for completeness.

Lemma 3.3. *Let $N = N^{\mathsf{T}} \in \mathbb{R}^{(p+m)\times(p+m)}$ be given and suppose $G \in \mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2e}^m \to \mathcal{L}_{2e}^p)$ is a finite-dimensional system with frequency response $\hat{G} \in \mathcal{L}_\infty^{p\times m}$. The following statements are equivalent.*

1. *The following frequency-domain inequality (FDI) holds for almost all $\omega \in \mathbb{R}$.*
$$\begin{bmatrix}\hat{G}(\omega)\\ I\end{bmatrix}^* N \begin{bmatrix}\hat{G}(\omega)\\ I\end{bmatrix} \succeq 0. \tag{3.17}$$
2. *Suppose G has a minimal realization (A, B, C, D). Then the following linear matrix inequality (LMI) has a solution $P = P^{\mathsf{T}}$. In the case $\mathcal{L}_2 = L_2$ (continuous time),*
$$\begin{bmatrix}A^{\mathsf{T}}P + PA & PB\\ B^{\mathsf{T}}P & 0\end{bmatrix} \preceq \begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^{\mathsf{T}} N \begin{bmatrix}C & D\\ 0 & I\end{bmatrix}, \tag{3.18}$$
and in the case $\mathcal{L}_2 = \ell_2$ (discrete time),
$$\begin{bmatrix}A^{\mathsf{T}}PA - P & A^{\mathsf{T}}PB\\ B^{\mathsf{T}}PA & B^{\mathsf{T}}PB\end{bmatrix} \preceq \begin{bmatrix}C & D\\ 0 & I\end{bmatrix}^{\mathsf{T}} N \begin{bmatrix}C & D\\ 0 & I\end{bmatrix}. \tag{3.19}$$

Proof. See for example [64]. ■

Remark 3.7. *Standard Lyapunov arguments may be applied to further refine Lemma 3.3. For example, If $N_{11} \preceq 0$ and G is stable, then we have $P \succeq 0$ in* (3.18) *or* (3.19).

3.4 Recovering Existing Results

Our results of Sections 3.2 and 3.3 can be used to recover a variety of existing robust stability results from the literature. To demonstrate this versatility, we begin with *necessity-preserving specializations*. These are choices that do not affect any of our proofs and thus yield equally strong results.

Table 3.2: Sufficient conditions for stability drawn from the literature. For each result, we show the choices of M and N and the condition $M+N \prec 0$ that allows Corollary 3.1 to recover the result. Under the assumptions of Theorem 3.4 these conditions become necessary as well, as described in Section 3.4. We use a positive feedback convention as in Fig. 3.1. To use the negative feedback convention instead, replace N by $\tilde{N}$ as described in Section 3.4.1.2.

Name of Theorem	$\boldsymbol{M}$	$\boldsymbol{N}$	$\boldsymbol{M+N \prec 0}$
Conic sector theorem [41, Thm. 2a, all three cases] or [51, Thm. 3.1, all parts of Case 1]	$\begin{bmatrix} \frac{-(a+\Delta)(b-\Delta)}{b-a-2\Delta} & \frac{-a-b}{2(b-a-2\Delta)} \\ \frac{-a-b}{2(b-a-2\Delta)} & \frac{-1}{b-a-2\Delta} \end{bmatrix}$	$\begin{bmatrix} \frac{ab}{b-a+2ab\delta} & \frac{a+b}{2(b-a+2ab\delta)} \\ \frac{a+b}{2(b-a+2ab\delta)} & \frac{(1+a\delta)(1-b\delta)}{b-a+2ab\delta} \end{bmatrix}$	$a < b$, and either $\delta = 0, \Delta > 0$ or $\delta > 0, \Delta = 0$.
Extended conic sector theorem [51, Thm. 3.1, all parts of Case 2]	$\begin{bmatrix} \frac{(a-\Delta)(b+\Delta)}{b-a+2\Delta} & \frac{a+b}{2(b-a+2\Delta)} \\ \frac{a+b}{2(b-a+2\Delta)} & \frac{1}{b-a+2\Delta} \end{bmatrix}$	$\begin{bmatrix} \frac{-ab}{b-a-2ab\delta} & \frac{-a-b}{2(b-a-2ab\delta)} \\ \frac{-a-b}{2(b-a-2ab\delta)} & \frac{-(1-a\delta)(1+b\delta)}{b-a-2ab\delta} \end{bmatrix}$	Same as above.
Extended passivity [3, Thm. 6.6.58]	$\begin{bmatrix} -\varepsilon_2 & \frac{1}{2} \\ \frac{1}{2} & -\delta_2 \end{bmatrix}$	$\begin{bmatrix} -\delta_1 & -\frac{1}{2} \\ -\frac{1}{2} & -\varepsilon_1 \end{bmatrix}$	$\delta_1 + \varepsilon_2 > 0$ and $\delta_2 + \varepsilon_1 > 0$.
Small gain theorem [36, Thm. 5.6]	$\begin{bmatrix} \gamma_2 & 0 \\ 0 & -1/\gamma_2 \end{bmatrix}$	$\begin{bmatrix} -1/\gamma_1 & 0 \\ 0 & \gamma_1 \end{bmatrix}$	$\gamma_1\gamma_2 < 1$.

3.4.1 Necessity-Preserving Specializations

3.4.1.1 Different Spaces

The most common spaces are $\mathcal{L}_{2e} = L_{2e}$ and $\mathcal{L}_{2e} = \ell_{2e}$ (continuous or discrete time, respectively). The results are essentially identical in these two cases. We saw in the proof of Theorem 3.4 that it is also possible to apply Theorem 3.1 directly to a space $\mathcal{L}_\infty$ of frequency responses.

3.4.1.2 Sign Conventions

Although we used the positive feedback sign convention in Fig. 3.1, using the negative feedback convention instead simply amounts to replacing N by $\tilde{N}$ in Theorem 3.3, Corollary 3.1, and Theorem 3.5, where

$$N := \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \quad \text{and} \quad \tilde{N} := \begin{bmatrix} N_{11} & -N_{12} \\ -N_{21} & N_{22} \end{bmatrix}.$$

Alternatively, we could replace M by $\tilde{M}$ (multiply the off-diagonal elements of M by -1). These changes amount to replacing G by $-G$ or Φ by $-\Phi$, respectively.

3.4.1.3 Strictness of Inequalities

Robust stability results generally involve non-intersecting sets (typically cones), so for the inequalities describing the admissible systems G and Φ, one will typically be strict and the other will be nonstrict. Our symmetric formulations using a coupling constraint $M + N \prec 0$ avoids the need to associate strictness with either G or Φ; both cases can be represented by suitable choice of M and N.

3.4.1.4 Different Cones

Different choices of the matrices M and N in our results allow the representation of different cones. For example, we can represent different flavors of passivity (input-strict passivity, output-strict passivity, extended passivity), small-gain results, the circle criterion, and other conic sectors that allow G or Φ to be unbounded/unstable.

To illustrate these various transformations, consider for example the classical passivity result by Vidyasagar, which is a sufficient-only result, and may be found in [3, Thm. 6.7.43].

Theorem 3.6 (Vidyasagar)**.** *Consider the system*

$$\begin{cases} e_1 = u_1 - y_2, \quad y_1 = Ge_1 \\ e_2 = u_2 + y_1, \quad y_2 = \Phi e_2 \end{cases}$$

Suppose there exist constants ε_1, ε_2, δ_1, δ_2 such that for all $\xi \in \ell_{2\mathrm{e}}$ and for all $T \geq 0$

$$\langle \xi,\, G\xi \rangle_T \geq \varepsilon_1 \|\xi\|_T^2 + \delta_1 \|G\xi\|_T^2, \tag{3.20a}$$

$$\langle \xi,\, \Phi\xi \rangle_T \geq \varepsilon_2 \|\xi\|_T^2 + \delta_2 \|\Phi\xi\|_T^2. \tag{3.20b}$$

Then the system is ℓ_2-stable if $\delta_1 + \varepsilon_2 > 0$ and $\delta_2 + \varepsilon_1 > 0$.

Theorem 3.6 uses a negative sign convention and is expressed in discrete time. To obtain a corresponding sufficient result, apply Corollary 3.1 with $\mathcal{L}_{2\mathrm{e}} = \ell_{2\mathrm{e}}$. Comparing (3.8) and (3.10) to (3.20), which yields the following $\tilde{N}$, N, and M.

$$\tilde{N} = \begin{bmatrix} -\delta_1 & \frac{1}{2} \\ \frac{1}{2} & -\varepsilon_1 \end{bmatrix}, \; N = \begin{bmatrix} -\delta_1 & -\frac{1}{2} \\ -\frac{1}{2} & -\varepsilon_1 \end{bmatrix}, \; M = \begin{bmatrix} -\varepsilon_2 & \frac{1}{2} \\ \frac{1}{2} & -\delta_2 \end{bmatrix}.$$

In Corollary 3.1, we require $M + N \prec 0$; thus $\delta_1 + \varepsilon_2 > 0$ and $\delta_2 + \varepsilon_1 > 0$, which recovers Theorem 3.6.

In Theorem 3.6 (and in Corollary 3.1), the systems G and Φ need not be linear. If we want a necessary version of this passivity result, we can assume $\mathcal{C}_G \subseteq \mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2\mathrm{e}}) \subseteq \mathcal{C}_\Phi$ and apply Theorem 3.4. In other words, we have necessity if G is LTI and we are certifying robustness with respect to a class $\mathcal{C}_\Phi$ that includes all LTI systems. This last requirement means that Φ may have *memory*.

Corollary 3.1 can also be specialized to recover the small-gain theorem [36, Thm. 5.6], extended conic sector theorem [51], circle criterion [65], and other versions of passivity such as Vidyasagar [3, Thm. 6.6.58] and Khong & van der Schaft [45]. We summarize these results in Table 3.2, along with the appropriate choice of M, N, and the associated condition $M + N \prec 0$.

3.4.2 Recovering Sufficient-Only Results

Many robust stability results in the literature (see Tables 3.1 and 3.2) are presented as being sufficient but *not* necessary. For those results involving *static multipliers* (passivity, small gain, conicity, etc.), we will argue that sufficiency follows from Theorem 3.1 or Corollary 3.1, while the lack of necessity follows from the necessary conditions described in Theorems 3.2, 3.4, and 3.5 not being met. In other words, we can view these sufficient results as direct consequences of their necessary-and-sufficient counterparts. We now discuss three ways in which this phenomenon can manifest itself in practice.

3.4.2.1 Alternative Inner Products

We showed that in Theorem 3.1, the partial converse (1) $\Longleftarrow$ (2) holds, and this fact is independent of the choice of semi-inner product. However, when we specialized to $\mathcal{L}_{2\mathrm{e}}$ in Section 3.3, we sought to prove input-output stability. Proving stability requires showing that certain properties of $\mathcal{L}_{2\mathrm{e}}$ signals imply boundedness in some other norm (e.g., the $\mathcal{L}_2$ norm). Our ability to do this *does* depend on the choice of semi-inner product used.

As an example, consider the space $\mathcal{V} = \mathcal{L}_{2\mathrm{e}}$ and define the *pointwise* semi-inner product as $\langle x,\, y \rangle_{p,T} := x(T)^* y(T)$. This may arise, for example, if G and Φ satisfy a stronger notion of *pointwise passivity* rather than the standard definition using $\langle \cdot,\, \cdot \rangle_T$. In this case, we can prove a result very similar to Corollary 3.1, except we can only prove the implications (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3). The partial converse (1) $\Longleftarrow$ (2) does not hold because having $\|y\| \le \gamma \|u\|$ for all $u \in \mathcal{L}_2$ *does not* imply that $|y(T)| \le \gamma |u(T)|$ for all $T \ge 0$.

3.4.2.2 Relaxed Constraints on Subsystems

A straightforward way to relax a necessary-and-sufficient result is to relax the conic constraints that characterize G and Φ. For example, consider Theorem 3.5. In Item (1), replace the condition $M + N \prec 0$ by $M + N \prec -\eta I$ for some $\eta > 0$. Equivalently, replace N by $\hat{N} \preceq \lambda N$ for some $\lambda > 0$ in (3.14). Naming the new condition (i′), we have that (i′) $\Longrightarrow$ (1). Likewise, define (ii′) to be the same as (2) except M is replaced by some $\hat{M} \preceq \mu M$ for some $\mu > 0$ in (3.15). Then, we have that (2) $\Longrightarrow$ (ii′). Putting these facts together, we obtain the implications: (i′) $\Longrightarrow$ (1) $\Longleftrightarrow$ (2) $\Longrightarrow$ (ii′). The implication (i′) $\Longrightarrow$ (ii′) cannot be reversed in general, so the necessary-and-sufficient condition has become sufficient-only.

Geometrically, constraints such as $\hat{M} \preceq \lambda M$ correspond to nested cones (the S-lemma). For example, consider the sets:

$$S := \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; \begin{bmatrix} x \\ y \end{bmatrix}^\mathsf{T} M \begin{bmatrix} x \\ y \end{bmatrix} \geq 0 \right\} \qquad \text{with } M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\hat{S} := \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; \begin{bmatrix} x \\ y \end{bmatrix}^\mathsf{T} \hat{M} \begin{bmatrix} x \\ y \end{bmatrix} \geq 0 \right\} \qquad \text{with } \hat{M} = \begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix}$$

The set S is the conic region between $y = -x$ and $y = x$, and $\hat{S}$ is the conic region between $y = 0$ and $y = x$. We have $\hat{S} \subseteq S$ and also $\hat{M} \preceq M$. Importantly, the implication is reversed when considering robust constraint satisfaction such as in (2). For example, given some property $\mathbb{P}$, we have:

$$\mathbb{P} \text{ holds for all } \Phi \in S \implies \mathbb{P} \text{ holds for all } \Phi \in \hat{S}.$$

We can also relax the constraints on G and Φ by applying Proposition 3.1. Namely, if we use $\mathcal{C}'_G \supseteq \mathcal{C}_G$ and $\mathcal{C}'_\Phi \subseteq \mathcal{C}_\Phi$ in any of our results, we will introduce conservatism.

3.4.2.3 Memoryless Nonlinearities

As mentioned in the previous item, choosing $\mathcal{C}'_\Phi \subseteq \mathcal{C}_\Phi$ will generally introduce conservatism. Perhaps the most famous such constraint is to restrict $\Phi : \mathcal{L}_{2\mathrm{e}} \to \mathcal{L}_{2\mathrm{e}}$ to be *memoryless*, which means that Φ operates pointwise in time. In this case, we can write $(\Phi x)(t) = \phi_t(x(t))$ for some functions ϕ_t. Note that when Φ is memoryless, we have:

$$\begin{aligned} & \left\langle \begin{bmatrix} x \\ \Phi x \end{bmatrix}, M \begin{bmatrix} x \\ \Phi x \end{bmatrix} \right\rangle_T \geq 0 \quad \text{for all } x \in \mathcal{L}_{2\mathrm{e}}, T \geq 0 \\ \iff & \left\langle \begin{bmatrix} x \\ \Phi x \end{bmatrix}, M \begin{bmatrix} x \\ \Phi x \end{bmatrix} \right\rangle_{p,T} \geq 0 \quad \text{for all } x \in \mathcal{L}_{2\mathrm{e}}, T \geq 0. \end{aligned}$$

where $\langle \cdot, \cdot \rangle_{p,T}$ is the pointwise inner product defined in Section 3.4.2.1. Indeed, the classical circle criterion is typically stated with a memoryless or static[8] nonlinearity Φ and a pointwise inner product characterizing the conic constraint. Finding a necessary and sufficient condition for robust stability under these assumptions remains an open problem [66].

Brockett's counterexample: In [10], a counterexample was presented that showed the standard circle criterion is only sufficient when Φ is memoryless. The closed-loop system for this example is described by the differential equation

$$\ddot{y} + 2\dot{y} + f(t)y = 0, \tag{3.21}$$

[8] The map $\Phi : \mathcal{L}_{2\mathrm{e}} \to \mathcal{L}_{2\mathrm{e}}$ is called *static* if it is both memoryless and time-invariant. A static Φ satisfies $(\Phi x)(t) = \phi(x(t))$ for some function ϕ.

where $f(t)$ satisfies $0 \leq f(t) \leq k$ for all t. In the context of Theorem 3.1, G is described by the transfer function $\frac{-1}{s(s+2)}$ and Φ is linear and memoryless, described by $(\Phi y)(t) = f(t)y(t)$.

The constraint on f corresponds to using $M = \left[\begin{smallmatrix} 0 & k \\ k & -2 \end{smallmatrix}\right]$. Applying Corollary 3.1, we seek to satisfy Item (1). Using $N = -M - \varepsilon I$ and applying Lemma 3.2, this amounts to certifying that for all $\omega \in \mathbb{R}$, we have

$$\begin{bmatrix} \hat{G}(\omega) \\ 1 \end{bmatrix}^* M \begin{bmatrix} \hat{G}(\omega) \\ 1 \end{bmatrix} < 0, \quad \text{where } \hat{G}(\omega) = \frac{-1}{j\omega(j\omega+2)}.$$

This condition simplifies to $k < 4+\omega^2$. Therefore, we conclude that we have robust stability of the closed-loop map (3.21) whenever $0 \leq k \leq 4$. It is explained in [10] that robust stability actually holds for $0 \leq k \leq 11.6$, therefore the circle criterion is sufficient-only.

The example above satisfies all the conditions of Theorem 3.5 except that $\mathcal{C}_\Phi \not\supseteq \mathscr{L}_{\mathrm{TI}}(\mathcal{L}_{2\mathrm{e}})$, since Φ is required to be memoryless and LTI systems are not memoryless in general.

If we allow Φ to be LTI and assume $k = 4 + \varepsilon$ for any $\varepsilon > 0$, then the frequency-domain condition is violated for any $0 < \omega_0 < \sqrt{\varepsilon}$. Following the construction described in the proof of Theorem 3.4, we can construct a Φ that is a static gain cascaded with a pure delay that depends on ω_0 and achieves arbitrarily large input-output gain. Such a Φ is *not* memoryless.

Therefore, the condition $0 \leq k \leq 4$ does render the circle criterion necessary and sufficient if Φ can have memory.

3.5 Conclusion

We studied robust stability results involving a plant G connected with a nonlinearity Φ belonging to a conic sector, e.g. passivity, small-gain, circle criterion, conicity, or extended conicity. Our goal was to distill the vast literature on this topic and state the most general and unified results possible.

Robust boundedness results are often stated in the form of *sufficient conditions.* Our first observation is that assumptions made in these results can always be relaxed in such a way that the sufficient conditions become necessary as well.

We distinguish between two types of necessity that are often confounded in the literature. In particular, when the sufficient conditions for robust boundedness are not met, we may seek:

1. *Worst-case signals* that yield an unbounded closed loop. This form of necessity *always* holds, even for nonlinear plants (Theorem 3.1 and Corollary 3.1).

2. *A worst-case nonlinearity* Φ that yields an unbounded closed loop. This form of necessity *always* holds in the general semi-inner product setting, and we show how to construct a *linear* worst-case Φ (Theorem 3.3). In the extended ($\mathcal{L}_{2\mathrm{e}}$) setting where G and Φ are *causal operators*, we show how to

construct a linear worst-case Φ when the plant is LTI (Theorem 3.4). Our constructed Φ consists of a static gain cascaded with a pure time delay.

Looking beyond the scope of this dissertation, it would be interesting to see if our semi-inner product framework could be used to recover results involving dynamic constraints (dissipativity, multiplier theory, integral quadratic constraints).

Our work also delineates (see Section 3.4) the ways in which necessity may be lost. This points to areas where the existing sufficient conditions could potentially be improved. Of particular note are problems where alternative inner products are used, or problems where the nonlinearity is memoryless.

3.6 Proofs

3.6.1 Proofs for Semi-Inner Product Spaces

This section contains proofs related to the partial necessity of Theorem 3.1 (Lemma 3.1) and the full necessity for the case where $\mathcal{C}_\Phi$ contains all linear relations (Theorem 3.3).

3.6.1.1 Proof of Lemma 3.1

Proof. Let $\gamma > 0$ be arbitrary. Here are the steps to the construction.

1. Since M is indefinite, we can write $M = P^* \left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right] P$ for some invertible $P \in \mathbb{F}^{2\times 2}$.

2. Define: $\tilde{\gamma} := \underline{\sigma}(P)^{-1}\left(\gamma\,\overline{\sigma}(P) + \overline{\sigma}\left(\left[\begin{smallmatrix}P_{12} & 0\\ 0 & P_{21}\end{smallmatrix}\right]\right)\right)$, where $\overline{\sigma}(\cdot)$ and $\underline{\sigma}(\cdot)$ denote the maximum and minimum singular value, respectively.

3. Choose any $\delta \in \mathbb{R}$ such that $0 < \delta < \frac{1}{\tilde{\gamma}+1}$. Rearranging this inequality, we obtain $\tilde{\gamma} < \frac{1-\delta}{\delta} < \frac{1}{\delta}$. Therefore,
$$(1-\delta)^2 - \tilde{\gamma}^2\delta^2 > 0 \quad \text{and} \quad 1 - \tilde{\gamma}^2\delta^2 > 0. \tag{3.22}$$

4. Choose any $\varepsilon \in \mathbb{R}$ such that $0 < \varepsilon < \frac{2\delta}{1+\delta^2}$. Rearranging this inequality, we obtain
$$\sqrt{\frac{1+\varepsilon}{1-\varepsilon}} < \frac{1+\delta}{1-\delta}. \tag{3.23}$$

5. Let $N = -M - \varepsilon P^* P$ and let ξ be such that
$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle < 0. \tag{3.24}$$

6. Define the signals $(\tilde{u}, \tilde{y}, \tilde{e})$ as follows:

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{e}_1 \end{bmatrix} := P \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, \tag{3.25}$$

$$\begin{bmatrix} \tilde{e}_2 \\ \tilde{y}_2 \end{bmatrix} := \begin{bmatrix} 1+\delta & 0 \\ 0 & 1-\delta \end{bmatrix} \begin{bmatrix} \tilde{y}_1 \\ \tilde{e}_1 \end{bmatrix}, \tag{3.26}$$

$$\begin{bmatrix} -\tilde{u}_2 \\ \tilde{u}_1 \end{bmatrix} := \begin{bmatrix} -\delta & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \tilde{y}_1 \\ \tilde{e}_1 \end{bmatrix}. \tag{3.27}$$

7. Finally, define the transformed signals:

$$\begin{bmatrix} y_1 \\ e_1 \end{bmatrix} := P^{-1} \begin{bmatrix} \tilde{y}_1 \\ \tilde{e}_1 \end{bmatrix} = \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, \tag{3.28}$$

$$\begin{bmatrix} e_2 \\ y_2 \end{bmatrix} := P^{-1} \begin{bmatrix} \tilde{e}_2 \\ \tilde{y}_2 \end{bmatrix}, \tag{3.29}$$

$$\begin{bmatrix} -u_2 \\ u_1 \end{bmatrix} := P^{-1} \begin{bmatrix} -\tilde{u}_2 \\ \tilde{u}_1 \end{bmatrix}. \tag{3.30}$$

Steps 1–7 provide the construction of (u, y, e). We now verify that this choice satisfies Items 1–3 of Lemma 3.1.

By adding and subtracting equations above, we obtain (3.25) = (3.26) + (3.27), and therefore (3.28) = (3.29) + (3.30). This immediately verifies that (3.1a), (3.1c), and (3.1d) are satisfied.

Based on how ξ is defined in (3.24), and substituting the choice of N from Step 5 and the factorization for M from Step 1, we obtain $(-1-\varepsilon)\|\tilde{y}_1\|^2 + (1-\varepsilon)\|\tilde{e}_1\|^2 < 0$. From this inequality, it is clear that $\|\tilde{y}_1\| \neq 0$. So we conclude that

$$\frac{\|\tilde{e}_1\|}{\|\tilde{y}_1\|} < \sqrt{\frac{1+\varepsilon}{1-\varepsilon}}. \tag{3.31}$$

Combining (3.31) and (3.23), we obtain $\frac{\|\tilde{e}_1\|}{\|\tilde{y}_1\|} < \frac{1+\delta}{1-\delta}$. Rearranging, we obtain $(1-\delta)\|\tilde{e}_1\| < (1+\delta)\|\tilde{y}_1\|$, which based on our definitions in Step 6, is equivalent to $\|\tilde{y}_2\| < \|\tilde{e}_2\|$. Rewriting as a quadratic form and converting coordinates, we can invoke the definitions in Steps 1 and 7 to obtain:

$$\left\langle \begin{bmatrix} \tilde{e}_2 \\ \tilde{y}_2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \tilde{e}_2 \\ \tilde{y}_2 \end{bmatrix} \right\rangle > 0 \iff \left\langle \begin{bmatrix} e_2 \\ y_2 \end{bmatrix}, M \begin{bmatrix} e_2 \\ y_2 \end{bmatrix} \right\rangle > 0.$$

This verifies Item 2 of Lemma 3.1. Based on the inequalities in (3.22) and the fact that $\|\tilde{y}_1\| > 0$ derived above, we have

$$\left(1 - \tilde{\gamma}^2\delta^2\right) \|\tilde{y}_1\|^2 + \left((1-\delta)^2 - \tilde{\gamma}^2\delta^2\right) \|\tilde{e}_1\|^2 > 0. \tag{3.32}$$

Applying the definitions from (3.26) and (3.27), Equation (3.32) is equivalent to:

$$\|\tilde{y}_1\|^2 + \|\tilde{y}_2\|^2 > \tilde{\gamma}^2 \left(\|\tilde{u}_1\|^2 + \|\tilde{u}_2\|^2\right),$$

or more compactly, $\|\tilde{y}\| > \tilde{\gamma}\|\tilde{u}\|$. We now apply the definitions (3.28)–(3.30) and the closed-loop equations (3.1a), (3.1c), (3.1d) to obtain a bound in terms of (u, y, e). For the upper bound,

$$\begin{aligned}\|\tilde{y}\| &= \left\| \begin{bmatrix} P_{11}y_1 + P_{12}e_1 \\ P_{21}e_2 + P_{22}y_2 \end{bmatrix} \right\| = \left\| \begin{bmatrix} P_{11}y_1 + P_{12}(y_2 + u_1) \\ P_{21}(y_1 + u_2) + P_{22}y_2 \end{bmatrix} \right\| \\ &= \left\| P \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} P_{12} & 0 \\ 0 & P_{21} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right\| \\ &\le \overline{\sigma}(P)\|y\| + \overline{\sigma}\left(\left[\begin{smallmatrix} P_{12} & 0 \\ 0 & P_{21} \end{smallmatrix}\right]\right)\|u\|.\end{aligned}$$

For the lower bound,

$$\tilde{\gamma}\|\tilde{u}\| = \tilde{\gamma}\left\| P \begin{bmatrix} -u_2 \\ u_1 \end{bmatrix} \right\| \ge \tilde{\gamma}\,\underline{\sigma}(P)\|u\|.$$

Combining the upper and lower bounds, we obtain:

$$\tilde{\gamma}\,\underline{\sigma}(P)\|u\| < \overline{\sigma}(P)\|y\| + \overline{\sigma}\left(\left[\begin{smallmatrix} P_{12} & 0 \\ 0 & P_{21} \end{smallmatrix}\right]\right)\|u\|.$$

Rearranging, we obtain

$$\underbrace{\overline{\sigma}(P)^{-1}\Big(\tilde{\gamma}\,\underline{\sigma}(P) - \overline{\sigma}\left(\left[\begin{smallmatrix} P_{12} & 0 \\ 0 & P_{21} \end{smallmatrix}\right]\right)\Big)}_{\gamma}\|u\| < \|y\|.$$

Based on how we defined $\tilde{\gamma}$ in Step 2, the above simplifies to $\|y\| > \gamma\|u\|$, which proves Item 3 from Lemma 3.1. ∎

3.6.1.2 Proof of Theorem 3.3

We begin by proving the following lemma, which states that a single pair of points satisfying a quadratic constraint can be extended to a linear relation that satisfies the quadratic constraint everywhere.

Lemma 3.4 (extension lemma)**.** *Let $\mathcal{V}$ be a semi-inner product space and let $M = M^* \in \mathbb{F}^{2\times 2}$ be indefinite. Suppose $e, y \in \mathcal{V}$ satisfy the inequality*

$$\left\langle \begin{bmatrix} e \\ y \end{bmatrix}, M \begin{bmatrix} e \\ y \end{bmatrix} \right\rangle \ge 0.$$

There exists $\Phi \in \mathscr{L}(\mathcal{V})$ such that:

1. *$(e, y) \in \Phi$.*

2. $\left\langle \begin{bmatrix} x \\ \Phi x \end{bmatrix}, M \begin{bmatrix} x \\ \Phi x \end{bmatrix} \right\rangle \geq 0$ *for all* $x \in \operatorname{dom}(\Phi)$.

Moreover, if $\|e\| > 0$*, we can construct* Φ *that is a* linear function*, with* $\operatorname{dom}(\Phi) = \mathcal{V}$*.*

Using Lemma 3.4, we can prove Theorem 3.3 by contradiction. Indeed, if Item (2) of Theorem 3.1 fails, then for any $\gamma > 0$, there exist $e_2, y_2 \in \mathcal{V}$ such that (A.2) and (3.1a), (3.1c), (3.1d) hold, with $\|y\| > \gamma\|u\|$. Applying Lemma 3.4 to the pair (e_2, y_2), we can produce $\Phi \in \mathscr{L}(\mathcal{V}) \subseteq \mathcal{C}_\Phi$ such that (3.5) holds, and thus (3.1b) holds, $(u, y) \in R_{uy}$, and therefore Item (3) of Theorem 3.1 fails, as required. All that remains is to prove Lemma 3.4.

Proof. We begin by considering some special cases.

Special case with $\|e\| = 0$ In this case, we must have $\langle e,\, y\rangle = 0$ by Cauchy–Schwarz. If $\|y\| = 0$, then define $\Phi = \{(z, x) \mid \|z\| = \|x\| = 0\}$. This is a degenerate case. If $\|y\| > 0$ instead, we have by assumption that

$$M_{22}\|y\|^2 = \left\langle \begin{bmatrix} e \\ y \end{bmatrix}, M \begin{bmatrix} e \\ y \end{bmatrix} \right\rangle \geq 0.$$

Therefore, $M_{22} \geq 0$. Define $\Phi = \{(z, x) \mid \|z\| = 0\}$. Roughly, Φ is the linear relation whose graph is a *vertical line*.

Special case with $\|e\| > 0$ **and** $\|y\| = 0$ As in the previous case, we must have $\langle e,\, y\rangle = 0$. By assumption,

$$M_{11}\|e\|^2 = \left\langle \begin{bmatrix} e \\ y \end{bmatrix}, M \begin{bmatrix} e \\ y \end{bmatrix} \right\rangle \geq 0.$$

So, $M_{11} \geq 0$. Define $\Phi x = 0$ and we have

$$\left\langle \begin{bmatrix} x \\ \Phi x \end{bmatrix}, M \begin{bmatrix} x \\ \Phi x \end{bmatrix} \right\rangle = M_{11}\|x\|^2 \geq 0 \quad \text{for all } x \in \mathcal{V}.$$

Henceforth, we will assume that $\|e\| > 0$ and $\|y\| > 0$. Define the normalized vectors $\hat{e} := \frac{e}{\|e\|}$ and $\hat{y} := \frac{y}{\|y\|}$. Also define the normalized inner product $\rho := \langle \hat{e}, \hat{y}\rangle$. Note that by Cauchy–Schwarz, we have $|\rho| \leq 1$.[9]

Special case: $|\rho| = 1$ Define $\Phi x = \rho\frac{\|y\|}{\|e\|}x$ and obtain:

$$\left\langle \begin{bmatrix} x \\ \Phi x \end{bmatrix}, M \begin{bmatrix} x \\ \Phi x \end{bmatrix} \right\rangle = \frac{\|x\|^2}{\|e\|^2} \left\langle \begin{bmatrix} e \\ y \end{bmatrix}, M \begin{bmatrix} e \\ y \end{bmatrix} \right\rangle \geq 0.$$

[9] Recall that in general, inner products are elements of $\mathbb{F}$, so ρ may be a complex number.

General case: $|\rho| < 1$ Since M is indefinite, there must exist some $\eta \in \mathbb{F}$ such that $\left[\begin{smallmatrix}1\\ \eta\end{smallmatrix}\right]^* M \left[\begin{smallmatrix}1\\ \eta\end{smallmatrix}\right] > 0$. For any $x \in \mathcal{V}$, we can write $x = x_{ey} + x_\perp$, where x_{ey} is a linear combination of $\hat{e}$ and $\hat{y}$ and $x_\perp$ is orthogonal to both $\hat{e}$ and $\hat{y}$. This can be computed via Gram–Schmidt:

$$\begin{aligned} x_{ey} &:= \left(\frac{\langle \hat{e}, x\rangle - \rho\langle \hat{y}, x\rangle}{1-|\rho|^2}\right)\hat{e} + \left(\frac{\langle \hat{y}, x\rangle - \bar{\rho}\langle \hat{e}, x\rangle}{1-|\rho|^2}\right)\hat{y}, \\ x_\perp &:= x - x_{ey}. \end{aligned}$$

We also have $\|x\|^2 = \|x_{ey}\|^2 + \|x_\perp\|^2$. Define the unit vectors

$$\hat{e}_\perp := \frac{\hat{y} - \rho\hat{e}}{\sqrt{1-|\rho|^2}} \quad \text{and} \quad \hat{y}_\perp := \frac{\bar{\rho}\hat{y} - \hat{e}}{\sqrt{1-|\rho|^2}}.$$

The vectors $\hat{e}_\perp$ and $\hat{y}_\perp$ are orthogonal to $\hat{e}$ and $\hat{y}$, respectively. Write $M_{12} = |M_{12}|e^{i\varphi}$ (polar decomposition). Since $M_{21} = \overline{M}_{12}$, we have the identity: $e^{-2i\varphi}M_{12} = M_{21}$.

Finally, define Φ as:

$$\Phi x := \frac{\|y\|}{\|e\|}\Big(\langle \hat{e}, x_{ey}\rangle\hat{y} + e^{-2i\varphi}\langle \hat{e}_\perp, x_{ey}\rangle\hat{y}_\perp\Big) + \eta\, x_\perp.$$

The function Φ is linear. The bracketed term lies in the span of $\hat{e}$ and $\hat{y}$ and performs an isometry that maps $\hat{e} \mapsto \hat{y}$, followed by the scaling $\frac{\|y\|}{\|e\|}$. This ensures that $\Phi e = y$. The remainder of Φx acts on the part of x orthogonal to the span of $\hat{e}$ and $\hat{y}$, and simply scales by η. One can readily check that $\|\Phi x_{ey}\| = \frac{\|y\|}{\|e\|}\|x_{ey}\|$. and $\operatorname{Re}\Big(M_{12}\langle x_{ey}, \Phi x_{ey}\rangle\Big) = \frac{\|y\|}{\|e\|}\|x_{ey}\|^2\operatorname{Re}(M_{12}\rho)$. Thus, $\begin{bmatrix} x \\ \Phi x\end{bmatrix} = \begin{bmatrix} x_{ey} \\ \Phi x_{ey}\end{bmatrix} + \begin{bmatrix} x_\perp \\ \eta x_\perp\end{bmatrix}$ and

$$\left\langle \begin{bmatrix} x \\ \Phi x\end{bmatrix}, M\begin{bmatrix} x \\ \Phi x\end{bmatrix}\right\rangle = \left\langle \begin{bmatrix} x_{ey} \\ \Phi x_{ey}\end{bmatrix}, M\begin{bmatrix} x_{ey} \\ \Phi x_{ey}\end{bmatrix}\right\rangle + \left\langle \begin{bmatrix} x_\perp \\ \eta x_\perp\end{bmatrix}, M\begin{bmatrix} x_\perp \\ \eta x_\perp\end{bmatrix}\right\rangle.$$

The first term simplifies to

$$\begin{aligned} &\left\langle \begin{bmatrix} x_{ey} \\ \Phi x_{ey}\end{bmatrix}, M\begin{bmatrix} x_{ey} \\ \Phi x_{ey}\end{bmatrix}\right\rangle \\ &= M_{12}\|x_{ey}\|^2 + 2\operatorname{Re}\Big(M_{12}\langle x_{ey}, \Phi x_{ey}\rangle\Big) + M_{22}\|\Phi x_{ey}\|^2 \\ &= \|x_{ey}\|^2\left(M_{11} + 2\operatorname{Re}(M_{12}\rho)\frac{\|y\|}{\|e\|} + M_{22}\frac{\|y\|^2}{\|e\|^2}\right) \\ &= \frac{\|x_{ey}\|^2}{\|e\|^2}\left(M_{11}\|e\|^2 + 2\operatorname{Re}\left(M_{12}\langle e, y\rangle\right) + M_{22}\|y\|^2\right) \\ &= \frac{\|x_{ey}\|^2}{\|e\|^2}\left\langle \begin{bmatrix} e \\ y\end{bmatrix}, M\begin{bmatrix} e \\ y\end{bmatrix}\right\rangle \geq 0. \end{aligned}$$

The second term simplifies to

$$\left\langle \begin{bmatrix} x_\perp \\ \eta x_\perp\end{bmatrix}, M\begin{bmatrix} x_\perp \\ \eta x_\perp\end{bmatrix}\right\rangle = \|x_\perp\|^2\begin{bmatrix} 1 \\ \eta\end{bmatrix}^* M\begin{bmatrix} 1 \\ \eta\end{bmatrix} \geq 0.$$

Therefore, we have $\left\langle \begin{bmatrix} x \\ \Phi x \end{bmatrix}, M \begin{bmatrix} x \\ \Phi x \end{bmatrix} \right\rangle \geq 0$, as required. ■

3.6.2 Proofs for Extended Spaces

First, we state a useful result that connects norms in $\mathcal{L}_2$ with truncated norms in $\mathcal{L}_{2\mathrm{e}}$.

Proposition 3.2. *Suppose $H : \mathcal{L}_{2e} \to \mathcal{L}_{2e}$ is causal, and for all $x \in \mathcal{L}_2$, we have $\|Hx\| \leq \gamma \|x\|$. Then for all $x \in \mathcal{L}_{2e}$ and $T \geq 0$, we have $\|Hx\|_T \leq \gamma \|x\|_T$.*

Proof. This result appears for example in [3, Lem. 6.2.11]. The proof is short so we reproduce it here. Let $x \in \mathcal{L}_{2\mathrm{e}}$, and then: $\|Hx\|_T = \|Hx_T\|_T \leq \|Hx_T\| \leq \gamma \|x_T\| = \gamma \|x\|_T$. ■

3.6.2.1 Proof of Corollary 3.1

Choose $\mathcal{V} = \mathcal{L}_{2\mathrm{e}}$ with inner product $\langle \cdot, \cdot \rangle_T$ and let $\mathcal{C}_G \subseteq \mathscr{F}(\mathcal{L}_{2\mathrm{e}})$ and $\mathcal{C}_\Phi \subseteq \mathscr{F}(\mathcal{L}_{2\mathrm{e}})$.

To prove (1) $\Longrightarrow$ (2), note that from the proof of Theorem 3.1, the gain γ only depends on the choice of M and N, and not on the choice of semi-inner product (choice of T). Likewise, fixing Φ and $u \in \mathcal{L}_2$ yields the same R_{uy} for all T. Therefore, we have $\|y\|_T \leq \gamma \|u\|_T$ for all $T \geq 0$ and γ, y are independent of T. Since $u \in \mathcal{L}_2$, letting $T \to \infty$ implies $y \in \mathcal{L}_2$ and we obtain (3.11).

To prove (1) $\Longleftarrow$ (2), suppose that for all (u, y, e) with $u \in \mathcal{L}_2$ satisfying (3.9), (3.1a), (3.1c), (3.1d), we have $\|y\| \leq \gamma \|u\|$. Since $\mathcal{C}_G$ and $\mathcal{C}_\Phi$ are subsets of $\mathscr{F}(\mathcal{L}_{2\mathrm{e}})$, then G and Φ are causal and so the closed-loop map R_{uy} is causal whenever the interconnection is well-posed [62, Prop. 1.2.14]. By Proposition 3.2, we have $\|y_T\| \leq \gamma \|u\|_T$. Apply Theorem 3.1 as before to obtain (3.8). From the way N is constructed in the proof of Theorem 3.1 (Step 5 in Section 3.6.1.1), N is independent of T and therefore the same N may be used for all T.

Proving (2) $\Longrightarrow$ (3) is similar to the corresponding proof in Theorem 3.1. Consider some $\Phi \in \mathcal{C}_\Phi \subseteq \mathscr{F}(\mathcal{L}_{2\mathrm{e}})$ for which (3.10) holds. Next, let (u, y, e) be the solution of (3.1). In particular, (3.1b) holds, so setting $\xi = e_2$ in (3.10), we obtain (3.9) and the rest of the proof follows as in Section 3.2.1. ■

3.6.2.2 Proof of Lemma 3.2

Define $G_s := (P_{11}G + P_{12})(P_{21}G + P_{22})^{-1}$. Suppose Item 1 holds. Substituting the factorization for N into (3.12), we obtain

$$\left\langle \begin{bmatrix} P_{11}G\xi + P_{12}\xi \\ P_{21}G\xi + P_{22}\xi \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{11}G\xi + P_{12}\xi \\ P_{21}G\xi + P_{22}\xi \end{bmatrix} \right\rangle_T \geq 0$$

$$\Longleftrightarrow \|P_{11}G\xi + P_{12}\xi\|_T \leq \|P_{21}G\xi + P_{22}\xi\|_T.$$

Let $\zeta = P_{21}G\xi + P_{22}\xi$. Eliminating ξ from the above, we obtain: $\|G_s\zeta\|_T \le \|\zeta\|_T$ Taking the limit $T \to \infty$, we conclude that G_s is bounded in the $\mathcal{L}_2$ norm.

Whenever $P_{21}G\xi + P_{22}\xi \in \mathcal{L}_2$, boundedness of G_s implies that $P_{11}G\xi + P_{12}\xi \in \mathcal{L}_2$. In such a case, we can take the limit $T \to \infty$ in (3.12) and $\langle\cdot,\,\cdot\rangle_T$ becomes $\langle\cdot,\,\cdot\rangle$, the inner product on $\mathcal{L}_2$. Applying Parseval's theorem yields

$$\left\langle \begin{bmatrix} P_{11}\hat{G} + P_{12} \\ P_{21}\hat{G} + P_{22} \end{bmatrix} \hat{\xi},\; \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{11}\hat{G} + P_{12} \\ P_{21}\hat{G} + P_{22} \end{bmatrix} \hat{\xi} \right\rangle \ge 0.$$

Due to our freedom in choosing $\hat{\xi}$, we conclude that

$$\begin{bmatrix} P_{11}\hat{G}(\omega) + P_{12} \\ P_{21}\hat{G}(\omega) + P_{22} \end{bmatrix}^* \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{11}\hat{G}(\omega) + P_{12} \\ P_{21}\hat{G}(\omega) + P_{22} \end{bmatrix} \ge 0$$

holds for almost all $\omega \in \mathbb{R}$. Substituting the factorization for N, we obtain Item 2. Conversely, we can begin from (3.13) and apply the steps in reverse together with Proposition 3.2 and we recover Item 1. ■

3.6.2.3 Proof of Theorem 3.4

We proceed by contradiction. Suppose Item (2) in Corollary 3.1 fails. Since (1) $\iff$ (2), this is equivalent to (1) failing. By Lemma 3.2, either $G_s := (P_{11}G + P_{12})(P_{21}G + P_{22})^{-1}$ does not map $\mathcal{L}_2 \to \mathcal{L}_2$, or there must exist $\omega_0 \in \mathbb{R}$ such that

$$\begin{bmatrix} \hat{G}(\omega_0) \\ 1 \end{bmatrix}^* N \begin{bmatrix} \hat{G}(\omega_0) \\ 1 \end{bmatrix} < 0. \tag{3.33}$$

If the former is true, choose the following static gain Φ and (frequency domain) input signals. If $P_{21} \ne 0$ and $P_{22} \ne 0$, let

$$\hat{u}_1 = -\tfrac{P_{11}}{P_{21}}\hat{\eta}, \qquad \hat{u}_2 = -\tfrac{P_{12}}{P_{21}}\hat{\eta}, \qquad \Phi = -\tfrac{P_{21}}{P_{22}}. \tag{3.34}$$

This is a valid choice of Φ because it satisfies (3.10) upon substituting the factorization for M. Substituting Φ, u_1, u_2 into the loop equations (3.1), we obtain $\hat{y}_2 = \hat{G}_s\hat{\eta}$. So the closed-loop map is not stable. If $P_{21} = 0$, we have $P = I$, so $G_s = G$. Then simply pick $\Phi = 0$ and the closed-loop map is unstable. If $P_{22} = 0$, use the choice (3.34) with $P_{22} = \varepsilon > 0$. Since $\hat{y}_2 \to \hat{G}_s\hat{\eta}$ as $\varepsilon \to 0$, then For ε sufficiently small, the map from $\hat{\eta}$ to $\hat{y}_2$ is unstable.

If (3.33) holds instead, then fix $\gamma > 0$ and follow the construction used in Lemma 3.1 to prove $\neg(1) \Longrightarrow \neg(2)$, except use the complex numbers $\xi \mapsto 1$ and $G\xi \mapsto \hat{G}(\omega_0)$ to seed the construction in (3.25). This results in $u_1, u_2, y_1, y_2, e_1, e_2 \in \mathbb{C}$ that satisfy (3.1a), (3.1c), (3.1d), $\|y\| > \gamma\|u\|$, and

$$\begin{bmatrix} e_2 \\ y_2 \end{bmatrix}^* M \begin{bmatrix} e_2 \\ y_2 \end{bmatrix} \ge 0. \tag{3.35}$$

Now view the constructed (y, u, e) as *phasors*. For example, y_1 is multiplied by $e^{j\omega_0 t}$ and represents a sinusoidal input with frequency ω_0 and magnitude and phase equal to those of y_1. The other signals in (y, u, e) are also multiplied by $e^{j\omega_0 t}$ and interpreted similarly.

The set of ω_0 that satisfy (3.33) is open, so we may pick ω_0 such that $e_2 \neq 0$. Choose $\Phi = \frac{y_2}{e_2} = re^{-j\theta}$ where $r > 0$ and $\theta \in [0, 2\pi)$ are the polar representation. So Φ is a static gain r cascaded with a pure delay of $\frac{\theta}{\omega_0}$. By construction, the signals (y, u, e) satisfy (3.1) and so $y = R_{uy}u$. Now there are two possible cases. If the closed-loop map R_{uy} is unstable, then we have shown that (3) fails, as required, for an unstable system cannot have a finite $\mathcal{L}_2$ gain. If the closed-loop map R_{uy} is instead stable, then since (3.35) holds for the phasors (u, y, e) and multiplying each phasor by $e^{j\omega_0 t}$ does not change the instantaneous value of (3.35), we have

$$\left\langle \begin{bmatrix} e_2 \\ y_2 \end{bmatrix}, M \begin{bmatrix} e_2 \\ y_2 \end{bmatrix} \right\rangle_T \geq 0 \quad \text{for all } T \geq 0$$

and $\|y\|_T > \gamma \|u\|_T$ for the time-domain sinusoids (u, y, e). Due to stability of R_{uy}, this sinusoidal fixed point of the dynamics is stable, and has gain $\gamma > 0$, which was arbitrarily chosen. It follows that (3) fails, as required. ■

Alternative proof. Consider the set $\mathcal{V} = \mathcal{L}_2(\mathbb{R})$ of frequency domain signals (either continuous or discrete time). Let

$$\mathcal{C}_G \subseteq \left\{ \mathcal{M}_{\hat{G}} \;\middle|\; \hat{G} \in \mathcal{L}_\infty \right\} \subseteq \mathcal{C}_\Phi,$$

where $\mathcal{M}_{\hat{G}}$ is the multiplication operator corresponding to the essentially bounded frequency response $\hat{G}$. Applying Theorem 3.1, Equation (A.1) becomes

$$\left\langle \begin{bmatrix} \hat{G}(\omega)\hat{\xi}(\omega) \\ \hat{\xi}(\omega) \end{bmatrix}, N \begin{bmatrix} \hat{G}(\omega)\hat{\xi}(\omega) \\ \hat{\xi}(\omega) \end{bmatrix} \right\rangle \geq 0 \quad \text{for all } \hat{\xi} \in \mathcal{L}_2(\mathbb{R}),$$

which is equivalent to

$$\begin{bmatrix} \hat{G}(\omega) \\ 1 \end{bmatrix}^* N \begin{bmatrix} \hat{G}(\omega) \\ 1 \end{bmatrix} \geq 0 \quad \text{for almost all } \omega \in \mathbb{R},$$

which is equivalent to (3.8) via Lemma 3.2. Applying Theorem 3.3, we conclude that a linear function Φ may be used to certify the necessary direction of Corollary 3.1 and the rest of the proof proceeds as above. ■

Chapter 4

A Robust Accelerated Optimization Algorithm for Strongly Convex Functions

This chapter proposes an accelerated first-order algorithm we call the Robust Momentum Method for optimizing smooth strongly convex functions. The algorithm has a single scalar parameter that can be tuned to trade off robustness to gradient noise versus worst-case convergence rate. At one extreme, the algorithm is faster than Nesterov's Fast Gradient Method by a constant factor but more fragile to noise. At the other extreme, the algorithm reduces to the Gradient Method and is very robust to noise. The algorithm design technique is inspired by methods from classical control theory and the resulting algorithm has a simple analytical form. Algorithm performance is verified on a series of numerical simulations in both noise-free and relative gradient noise cases.

4.1 Introduction

Consider the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) \tag{4.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth and m-strongly convex. The strong convexity of f guarantees that there exists a unique minimizer $x_\star$ satisfying $\nabla f(x_\star) = 0$. First-order methods are widely used for solving (4.1) when the Hessian is prohibitively expensive to compute, e.g., when the problem dimension is large. A simple first-order algorithm for solving (4.1) is the Gradient Method (GM),

$$x_{k+1} = x_k - \alpha\nabla f(x_k), \qquad x_0 \in \mathbb{R}^n.$$

For smooth and strongly convex f, the GM with a well-chosen stepsize converges linearly to the optimizer [34]. That is, for some $c \geq 0$ and $\rho \in [0, 1)$, we have

$$\|x_k - x_*\| \leq c\,\rho^k \quad \text{for all } k \geq 0.$$

For example, the standard choice $\alpha = 1/L$ leads to a linear rate $\rho = 1 - \frac{m}{L}$, while the choice $\alpha = \frac{2}{L+m}$ results in the improved linear rate $\rho = \frac{L-m}{L+m}$.

The issue with the Gradient Method, however, is that the convergence rate is slow, especially for ill-conditioned problems where the ratio $\frac{L}{m}$ is large. A common method of accelerating convergence is to use *momentum.* A well-established momentum algorithm for smooth and strongly convex f is Nesterov's Fast Gradient Method[1], (FGM) [33] described by the iteration $x_0, x_{-1} \in \mathbb{R}^n$

$$\begin{aligned} x_{k+1} &= y_k - \alpha \nabla f(y_k), \\ y_k &= x_k + \beta(x_k - x_{k-1}). \end{aligned}$$

The FGM tuned with $\alpha = \frac{1}{L}$ and $\beta = \frac{\sqrt{L}-\sqrt{m}}{\sqrt{L}+\sqrt{m}}$ converges with rate $\rho^2 < 1 - \sqrt{m/L}$, which is faster than the GM rate[2]. The rate can be improved to $\rho = 1 - \sqrt{m/L}$ using an accelerated algorithm called the Triple Momentum Method [11]. This is the fastest known worst-case convergence rate for this class of problems. Robustness issues arise naturally in many optimization problems. For example, achieving the above rates associated with each first-order method requires knowledge of L and m, which may not be accurately accessible in practice. In addition, the gradient evaluation can be inexact for certain applications [67–69]. These issues motivate the need for accelerated first-order methods that are robust to underlying design assumptions.

As observed in [32, §5.2], optimization algorithm design involves a trade-off between performance and robustness. For example, consider step-size tuning for the GM. Using $\alpha = \frac{2}{L+m}$ optimizes the convergence rate, but makes the algorithm fragile to gradient noise. The more conservative choice $\alpha = \frac{1}{L}$ results in slower convergence, but more robustness to noise. This is consistent with the intuition that a smaller step-size can improve the algorithm's robustness at the price of degrading its performance. For momentum methods, exploiting the trade-off between performance and robustness is less straightforward, since one has to tune multiple algorithm parameters in a coupled manner to achieve acceleration. This tradeoff is exploited in [70] for first-order methods applied to smooth convex problems. In this work, we design a first-order method that exploits the trade-off between robustness and performance for smooth strongly convex problems.

4.2 Main Result

4.2.1 Robust Momentum Method

Our proposed algorithm is parameterized by a scalar ρ that represents the worst-case convergence rate of the algorithm in the noise-free case. Specifically, the iteration is governed by the following recursion with

[1] Also called Neterov's accelerated gradient method.

[2] A numerical study in [32] revealed that the standard rate bound for FGM derived in [33] is conservative. Nevertheless, the bound has a simple algebraic form and is asymptotically tight.

arbitrary initialization $x_0, x_{-1} \in \mathbb{R}^n$

$$x_{k+1} = x_k + \beta(x_k - x_{k-1}) - \alpha \nabla f(y_k), \tag{4.2a}$$

$$y_k = x_k + \gamma(x_k - x_{k-1}). \tag{4.2b}$$

where α, β, and γ depend directly on the parameter ρ as

$$\alpha = \frac{\kappa(1-\rho)^2(1+\rho)}{L}, \qquad \beta = \frac{\kappa\rho^3}{\kappa - 1}, \qquad \gamma = \frac{\rho^3}{(\kappa-1)(1-\rho)^2(1+\rho)}. \tag{4.3}$$

We now state the key convergence property of the Robust Momentum Method in the noise-free case.

Theorem 4.1. *Suppose $f \in \mathcal{F}(m, L)$ with $0 < m \le L$ and let $x_\star$ be the unique minimizer of f. Given the parameter $\rho \in [1 - 1/\sqrt{\kappa},\, 1 - 1/\kappa]$, the Robust Momentum Method* (4.2) *with parameter tuning* (4.3) *satisfies the bound*

$$\|x_k - x_\star\| \le c\,\rho^k \qquad \textit{for } k \ge 1 \tag{4.4}$$

where $c > 0$ is a constant that does not depend on k.

The proof of Theorem 4.1 is provided in Section 4.2.2. Theorem 4.1 states that ρ directly controls the worst-case convergence rate of the Robust Momentum Method. We will see in Section 4.3 that although increasing ρ makes the algorithm slower, it also makes it more robust to gradient noise. In particular,

- The minimum value is $\rho = 1 - 1/\sqrt{\kappa}$. This is the fastest achievable convergence rate and also leads to the most fragile algorithm. This choice recovers the Triple Momentum Method [11].

- The maximum value is $\rho = 1 - 1/\kappa$. This is the slowest achievable convergence rate and also leads to the most robust algorithm. This choice recovers the Gradient Method with stepsize $\alpha = 1/L$.

To see why this last case reduces to the Gradient Method, substitute $\rho = 1 - 1/\kappa$ into (4.2) and (4.3). Then, (4.2a) reduces to $y_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$.

4.2.2 Convergence Rate Proof

In this section, we derive a proof for Theorem 4.1. The approach that follows is similar to the one used in [32], with one important difference. In addition to proving a rate bound as in [32], we also derive a Lyapunov function that yields intuition for the algorithm's behavior and robustness properties.

Proposition 4.1 (Co-coercivity)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable. Further suppose f is L-smooth. Then for all $x, y \in \mathbb{R}^n$,*

$$f(y) \ge f(x) + \nabla f(x)^\mathsf{T}(y - x) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2.$$

The following lemma proves a key property of strongly convex functions. Parts of this result appear in [32] and we repeat them here for completeness.

Lemma 4.1. *Suppose $f \in \mathcal{F}(m,L)$. Let $x_\star$ be the unique minimizer of f (i.e., $\nabla f(x_\star) = 0$). Define the function $g(x) := f(x) - f(x_\star) - \frac{m}{2}\|x - x_\star\|^2$. Given any sequence of points $\{y_k\} \subseteq \mathbb{R}^n$,*

1. *If we define $q_k := (L-m)g(y_k) - \frac{1}{2}\|\nabla g(y_k)\|^2$, then*
$$q_k \geq 0 \quad \textit{for all } k.$$
2. *If we define $u_k := \nabla f(y_k)$ and $\tilde{y}_k := y_k - x_\star$, then*
$$(u_k - m\tilde{y}_k)^\mathsf{T}(L\tilde{y}_k - u_k) \geq q_k \quad \textit{for all } k.$$
3. *Using the same definitions as above, the following inequality holds for any $0 \leq \rho \leq 1$,*
$$(u_k - m\tilde{y}_k)^\mathsf{T}\big(L(\tilde{y}_k - \rho^2\tilde{y}_{k-1}) - (u_k - \rho^2 u_{k-1})\big) \geq q_k - \rho^2 q_{k-1} \quad \textit{for all } k.$$

Proof. By the definition of strong convexity, g is convex and $(L-m)$-smooth. Moreover, $g(y) \geq g(x_\star) = 0$ for all $y \in \mathbb{R}^n$. Item 1 follows from applying Proposition 4.1 with $(f,x,y) \mapsto (g, x_\star, y_k)$. For Item 2, note that $u_k = \nabla f(y_k) = \nabla g(y_k) + m\tilde{y}_k$. We have

$$\begin{aligned}(u_k - m\tilde{y}_k)^\mathsf{T}(L\tilde{y}_k - u_k) &= \nabla g(y_k)^\mathsf{T}\big((L-m)\tilde{y}_k - \nabla g(y_k)\big)\\ &\geq (L-m)g(y_k) - \tfrac{1}{2}\|\nabla g(y_k)\|^2\\ &= q_k\end{aligned}$$

where the inequality follows from applying Proposition 4.1 with $(f,x,y) \mapsto (g, y_k, x_\star)$. To prove Item 3, begin with the case $\rho = 1$. Using a similar argument to the one used to prove Item 2,

$$\begin{aligned}&(u_k - m\tilde{y}_k)^\mathsf{T}\big(L(\tilde{y}_k - \tilde{y}_{k-1}) - (u_k - u_{k-1})\big)\\ &= \nabla g(y_k)^\mathsf{T}\big((L-m)(\tilde{y}_k - \tilde{y}_{k-1}) - (\nabla g(y_k) - \nabla g(y_{k-1}))\big)\\ &\geq q_k - q_{k-1}\end{aligned}$$

where the inequality follows from applying Proposition 4.1 with $(f,x,y) \mapsto (g, y_k, y_{k-1})$. By combining the two previous results, we have

$$\begin{aligned}&(u_k - m\tilde{y}_k)^\mathsf{T}\big(L(\tilde{y}_k - \rho^2\tilde{y}_{k-1}) - (u_k - \rho^2 u_{k-1})\big)\\ &= (1-\rho^2)(u_k - m\tilde{y}_k)^\mathsf{T}\big(L\tilde{y}_k - u_k\big) + \rho^2(u_k - m\tilde{y}_k)^\mathsf{T}\big(L(\tilde{y}_k - \tilde{y}_{k-1}) - (u_k - u_{k-1})\big)\\ &\geq (1-\rho^2)q_k + \rho^2(q_k - q_{k-1})\\ &= q_k - \rho^2 q_{k-1}\end{aligned}$$

and this completes the proof of Item 3. ■

Our next lemma provides a key algebraic property of the Robust Momentum Method (4.2). This result makes no assumptions about f.

Lemma 4.2. *Suppose $\{u_k, x_k, y_k\}$ is any sequence of vectors satisfying the constraints*

$$\begin{bmatrix} x_{k+1} \\ y_k \end{bmatrix} = \begin{bmatrix} 1+\beta & -\beta & -\alpha \\ 1+\gamma & -\gamma & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \\ u_k \end{bmatrix} \quad \text{for } k \geq 0 \tag{4.5}$$

where (α, β, γ) are given by (4.3), and thus depend on the parameters $0 < m \leq L$, $\kappa := L/m$, and $\rho \in (0,1)$. Define $z_k := (1-\rho^2)^{-1}\left(x_k - \rho^2 x_{k-1}\right)$ for $k \geq 0$. Then the following algebraic identity holds for $k \geq 1$,

$$(u_k - my_k)^\mathsf{T}\left(L(y_k - \rho^2 y_{k-1}) - (u_k - \rho^2 u_{k-1})\right) + \lambda\left(\|z_{k+1}\|^2 - \rho^2\|z_k\|^2\right) + \nu\|u_k - my_k\|^2 = 0 \tag{4.6}$$

where the constants λ and ν are defined as

$$\lambda := \frac{m^2\left(\kappa - \kappa\rho^2 - 1\right)}{2\rho(1-\rho)} \quad \text{and} \tag{4.7}$$

$$\nu := \frac{(1+\rho)\left(1 - \kappa + 2\kappa\rho - \kappa\rho^2\right)}{2\rho}. \tag{4.8}$$

Proof. The algebraic identity may be verified by direct substitution of (4.3), (4.5), (4.7), and (4.8) into (4.6). Specifically, the constraints (4.5) allow us to express z_{k+1}, z_k, y_k, y_{k-1}, u_k, and u_{k-1} as linear functions of x_k, x_{k-1}, x_{k-2}, and u_k. Upon doing so, the resulting expression becomes identically zero. To express u_{k-1} as required, rearrange the first equation of (4.5) to obtain the expression $u_{k-1} = \alpha^{-1}((1+\beta)x_{k-1} - \beta x_{k-2} - x_k)$. ■

The algebraic identity (4.6) has three main terms. We will see how each serves a role in explaining the convergence and robustness properties of our algorithm. We are now ready to prove Theorem 4.1.

Proof of Theorem 4.1. Choose x_0 and x_{-1} arbitrarily and consider the sequence $\{u_k, x_k, y_k, z_k\}$ defined by setting $u_k := \nabla f(y_k)$ and propagating for all $k \geq 0$ using (4.5). This sequence is precisely a trajectory of our algorithm. Let $x_\star$ be the unique minimizer of f. Define the shifted sequences $\tilde{x}_k := x_k - x_\star$, $\tilde{y}_k := y_k - x_\star$, and $\tilde{z}_k := z_k - x_\star$ where z_k is defined in Lemma 4.2. Note that the constraints (4.5) still hold when we use the shifted sequence $\{u_k, \tilde{x}_k, \tilde{y}_k, \tilde{z}_k\}$. Applying Lemma 4.2 with Item 3 of Lemma 4.1, we conclude that for $k \geq 1$,

$$\lambda(\|\tilde{z}_{k+1}\|^2 - \rho^2\|\tilde{z}_k\|^2) + (q_k - \rho^2 q_{k-1}) + \nu\,\|u_k - m\tilde{y}_k\|^2 \leq 0, \tag{4.9}$$

where λ and ν are defined in (4.7)–(4.8). When $1 - 1/\sqrt{\kappa} \leq \rho \leq 1 - 1/\kappa$, we have $mL \geq \lambda \geq \frac{1}{2}mL$ and $0 \leq \nu \leq 1 - \frac{1}{2\kappa}$. As we increase ρ, the parameter λ decreases monotonically while ν increases monotonically.

Define the sequence $\{V_k\}$ by $V_k := \lambda\|\tilde{z}_k\|^2 + q_{k-1}$. If we choose ρ in the interval specified above, then $\nu \geq 0$ and $\lambda > 0$. Since $q_k \geq 0$, V_k can serve as a Lyapunov function. In particular, it follows from (4.9) that

$$V_{k+1} \leq \rho^2 \, V_k \qquad \text{for } k \geq 1. \tag{4.10}$$

Iterating this relationship, we find that $V_{k+1} \leq \rho^{2k} \, V_1$. The reason we do not iterate down to zero is because V_k is not defined at $k = 0$. Substituting the definitions and simplifying, we obtain the bound

$$\|\tilde{z}_{k+1}\| \leq \rho^k \sqrt{\|\tilde{z}_1\|^2 + \lambda^{-1} q_0} \qquad \text{for } k \geq 1. \tag{4.11}$$

The bound (4.11) therefore captures two effects. As we increase ρ, the linear rate ρ^k becomes slower and the constant factor in the rate bound also grows.

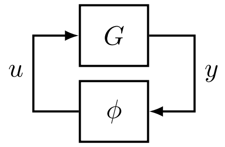
Next, we show that $\{\tilde{x}_k\}$ goes to zero at the same rate ρ^k, but with different constant factors. Note that because $\tilde{z}_k = (1-\rho^2)^{-1}\left(\tilde{x}_k - \rho^2 \tilde{x}_{k-1}\right)$, we can form the telescoping sum

$$\tilde{x}_k = \rho^{2(k-1)} \tilde{x}_{-1} + (1-\rho^2) \sum_{t=0}^{k-1} \rho^{2(k-t)} \tilde{z}_t \quad \text{for } k \geq 0. \tag{4.12}$$

Taking the norm of both sides of (4.12), applying the triangle inequality, and substituting (4.11), we obtain a geometric series. Upon simplification, we find that $\|\tilde{x}_k\|$ is bounded above by a constant times ρ^k, as required.

4.3 Control Design Interpretations

In this section, we cast the problem of algorithm analysis as a robust control problem. Specifically, we can view the problem of algorithm analysis as being equivalent to solving a Lur'e problem [1]. The Lur'e setup is illustrated in Figure 4.1, where a linear dynamical system G (4.13) is in feedback with a static nonlinearity ϕ.



$$\xi_{k+1} = A\xi_k + Bu_k, \tag{4.13a}$$

$$y_k = C\xi_k, \tag{4.13b}$$

$$u_k = \phi(y_k). \tag{4.13c}$$

Figure 4.1: Feedback interconnection of a linear system G with a troublesome (nonlinear or uncertain) component ϕ. We use the positive feedback convention in this block diagram.

The Robust Momentum Method (as well as the Fast Gradient Method and ordinary Gradient Method) can be written in this way by setting $\phi = \nabla f$ and choosing A, B, and C appropriately. For example, the

Robust Momentum Method (4.2) is given by

$$A = \begin{bmatrix} 1+\beta & -\beta \\ 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} -\alpha \\ 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 1+\gamma & -\gamma \end{bmatrix}.$$

Here, we shifted all signals so they are measured relative to the steady-state value $x_\star$ and therefore assumed that $\nabla f(0) = 0$. We also assumed without loss of generality that u_k and y_k are scalars. This interpretation was used in [32, 71, 72] to provide a unified analysis framework.

Traditionally, Lur'e systems were analyzed in the frequency domain rather than the time domain. For the case of the Robust Momentum Method, the (discrete-time) transfer function of the linear block is given by

$$G(z) = -\alpha \frac{(1+\gamma)z - \gamma}{(z-1)(z-\beta)}. \tag{4.14}$$

It was observed in Section 4.2.1 that the Robust Momentum Method becomes the Gradient Method if $\rho = 1 - 1/\kappa$. This fact can be directly verified using the transfer function. Substituting this ρ and the parameter values (4.3) into (4.14), there is a pole-zero cancellation and we obtain $G(z) = \frac{-1}{L(z-1)}$, which is the transfer function for the Gradient Method with stepsize $\alpha = \frac{1}{L}$.

Frequency-domain condition. Continuing with the frequency-domain interpretation, Lur'e systems can be analyzed using the formalism of Integral Quadratic Constraints (IQCs) [13]. To this end, the nonlinearity is characterized by a quadratic inequality that holds between its input and output

$$\int_{|z|=1} \begin{bmatrix} \hat{y}(z) \\ \hat{u}(z) \end{bmatrix}^* \Pi(z) \begin{bmatrix} \hat{y}(z) \\ \hat{u}(z) \end{bmatrix} \mathrm{d}z \geq 0$$

where $\hat{y}$ and $\hat{u}$ are the z-transforms of $\{y_k\}$ and $\{u_k\}$, respectively, and $\Pi(z)$ is a para-Hermitian matrix. For convenience, we use a loop-shifting transformation to move the nonlinearity $\phi = \nabla f$ from the sector (m, L) to the sector $(0, \kappa - 1)$. We also scale the frequency variable z by a factor of ρ so that we can reduce the problem of certifying exponential stability (finding a linear rate) to that of certifying BIBO stability. This procedure is described in [73].

The nonlinearity of interest is sector-bounded and slope-restricted because it is the gradient of a function $g \in \mathcal{F}(0, \kappa - 1)$. We may therefore represent the nonlinearity with a Zames–Falb IQC as in [73], leading to

$$\Pi(z) := \begin{bmatrix} 0 & (\kappa-1)(1-\rho^2\bar{z}^{-1}) \\ (\kappa-1)(1-\rho^2 z^{-1}) & -2+\rho^2(z^{-1}+\bar{z}^{-1}) \end{bmatrix}.$$

The transformed transfer function is

$$\tilde{G}(z) = \frac{-\alpha m(1+\gamma)z + \alpha m \gamma}{z^2 - (1+\beta-\alpha m(1+\gamma))z + \beta - \alpha m \gamma}. \tag{4.15}$$

To certify stability of the feedback interconnection, we must have $\tilde{G}(\rho z)$ stable and for all $|z| = 1$,

$$\mathrm{Re}\Big((1-\rho z^{-1})\big((\kappa-1)\tilde{G}(\rho z)-1\big)\Big) < 0. \tag{4.16}$$

Equation (4.16) has a graphical interpretation; that the Nyquist plot of $F(z) := (1-\rho z^{-1})\big((\kappa-1)\tilde{G}(\rho z)-1\big)$ should lie entirely in the left half-plane.

Graphical design for robustness. The frequency-domain condition (4.16) can provide useful intuition for the design of robust accelerated optimization methods. We can visualize different algorithms by choosing the parameters α, β, γ appropriately in (4.15).

In Figure 4.2 (left panel), we show the Nyquist plot for the Gradient Method using the sector IQC [32,73]. To this effect, we set $\beta = \gamma = 0$ and use either $\alpha = \frac{2}{L+m}$ or $\alpha = \frac{1}{L}$. As we increase ρ, the Nyquist plots become ellipses in the left half-plane. At the fastest certifiable rate (smallest ρ), the plots become vertical lines. When $\alpha = \frac{2}{L+m}$, the vertical line coincides with the imaginary axis, whereas when $\alpha = \frac{1}{L}$, the vertical line is shifted left. This result confirms our intuition that since the imaginary axis is the stability boundary, *robust* stability is achieved as the Nyquist contour moves further left, away from the boundary.

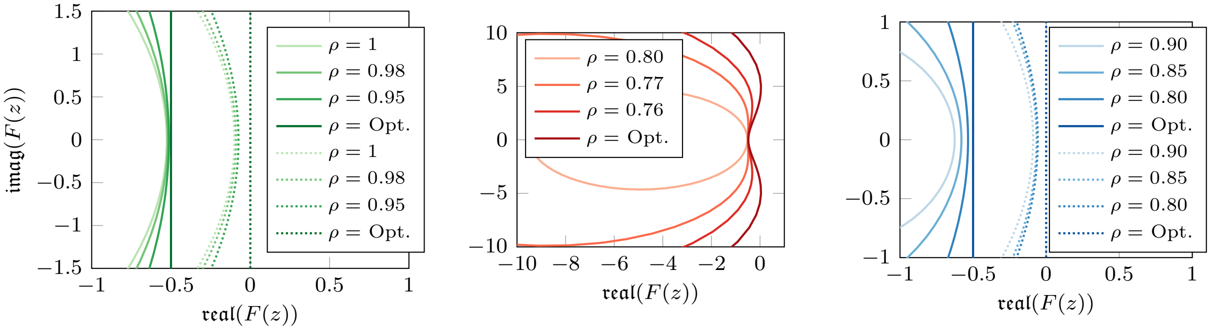
The Robust Momentum Method (4.2) was designed such that the Nyquist diagram forms a vertical line passing through the point $(-\nu, 0)$. In other words, we solved for (α, β, γ) such that (4.16) holds with the right-hand side replaced by $-\nu$. Constraining the Nyquist plot as such directly leads to the choice (4.3) with ν related to ρ via (4.8). In Figure 4.2 (right panel), we show the Nyquist plot for the Robust Momentum Method using the Zames–Falb IQC (for $\nu = 0$ and $\nu = \frac{1}{2}$). We also show Nyquist plots that certify a convergence rate of ρ that is larger than the corresponding algorithm parameter. This leads to ellipses as with the Gradient Method. Note that although the RMM and GM plots look similar, the RMM ρ-values are generally smaller due to acceleration. In contrast, the FGM (center panel) does not produce a vertical line in the Nyquist plot but still touches the stability boundary at the optimal ρ.

Further robustness interpretations. The parameter ν can be interpreted as the *input feed-forward passivity index* (IFP) [74], which is a measure of the shortage or excess of passivity of the system $F(z)$ defined above. In the frequency domain, the discrete-time definition of the IFP index is given by[3]

$$\nu(F(z)) := -\tfrac{1}{2} \max_{|z|=1} \lambda_{\max}\big(F(z) + F(z)^*\big), \tag{4.17}$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue and F^* is the conjugate transpose of F. For the SISO case, (4.17) reduces to $\nu = -\max_{|z|=1} \mathfrak{real}(F(z))$, which is the shortest distance between each curve and the imaginary axis in Figure 4.2.

[3] Most sources use a negative feedback convention. The definition we give in (4.17) uses the positive feedback convention.



(a) Gradient Method

(b) Fast Gradient Method

(c) Robust Momentum Method

Figure 4.2: Frequency-domain plots of various algorithms for $\kappa = 10$ and different values of the convergence rate ρ. The system is stable if the entire curve lies in the left half-plane. **(a)** Gradient Method for $\alpha = 1/L$ (solid) and $\alpha = 2/(L+m)$ (dashed). The latter is right on the stability boundary while the former is shifted left (more robust). **(b)** Fast Gradient Method. **(c)** Robust Momentum Method for $\nu = 1/2$ (solid) and $\nu = 0$ (dashed). Again, the latter is right on the stability boundary while the former is shifted left (more robust).

We can also interpret ν as a robustness margin in the time domain using the Lyapunov function defined in (4.8). In the proof of Theorem 4.1, when we substitute the definition for V_k into (4.9), we obtain

$$V_{k+1} \le \rho^2 \, V_k - \nu \, \|\nabla g(y_k)\|^2.$$

Proving the desired rate bound only requires (4.10) to hold, so the term $\nu \, \|\nabla g(y_k)\|^2$ can be interpreted as an additional margin that ensures the inequality $V_{k+1} \le \rho^2 V_k$ will hold even if underlying assumptions such as exactness in gradient evaluations or accurate knowledge of L and m are violated. As we increase ρ, the linear rate becomes slower, but ν also increases via (4.8), which serves to increase the robustness margin in the inequality (4.10).

4.4 Robustness to Gradient Noise

The Robust Momentum Method has a single parameter, which can be used to tune the performance. In this section, we provide both simulations and numerical rate analyses to verify the performance of the algorithm when the gradient is subject to relative deterministic noise [34]. Specifically, we will suppose that instead of measuring the gradient $\nabla f(y_k)$, we measure $u_k = \nabla f(y_k) + r_k$ where $r_k \in \mathbb{R}^n$ satisfies $\|r_k\| \le \delta \, \|\nabla f(y_k)\|$. For a given fixed $\delta \ge 0$, we will bound the worst-case performance of the algorithm over all $f \in \mathcal{F}(m, L)$ and feasible $\{r_k\}$.

Numerical rate analysis. To find the worst-case performance, we adopt the methodology from [32, Eq. 5.1]. There, the authors formulate a linear matrix inequality parameterized by $\hat{\rho}$ and δ whose feasibility provides a sufficient condition for convergence with linear rate $\hat{\rho}$.

In Figure 4.3, we plot the computed convergence rate as a function of noise strength δ for the Gradient Method, Fast Gradient Method, and Robust Momentum Method. Note that the worst-case rate in closed form for the Gradient Method is given in [75, 76].

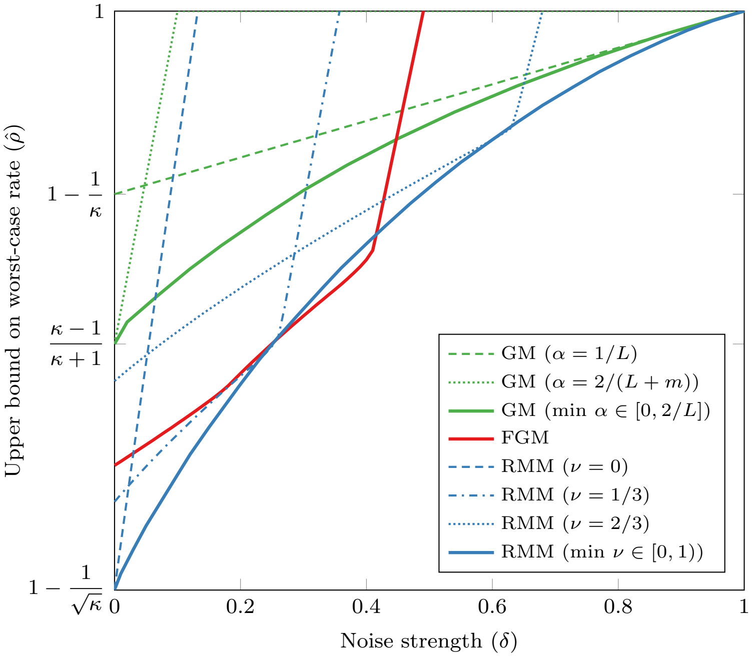
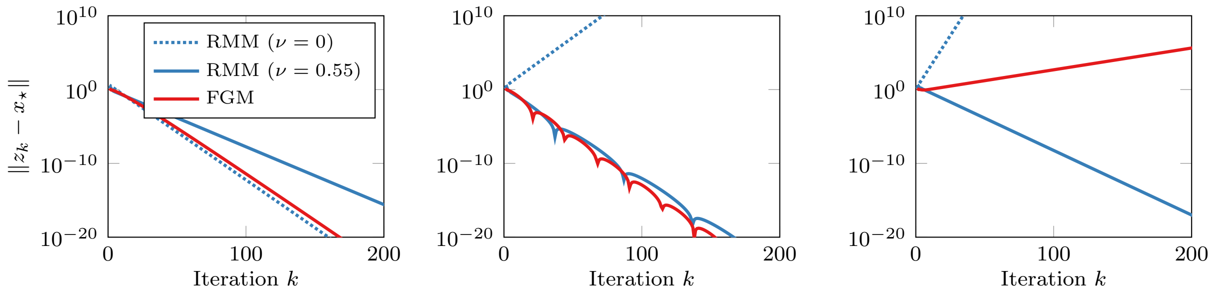


Figure 4.3: Upper bound on the worst-case linear convergence rate as a function of the noise level δ for $\kappa = 10$ (the figure looks similar for other choices of κ). We used a relative noise model, where the measured gradient u_k satisfies $\|u_k - \nabla f(y_k)\| \le \delta \|\nabla f(y_k)\|$ for the Gradient Method (GM), Fast Gradient Method (FGM), and Robust Momentum Method (RMM). By tuning the parameter ν, the RMM trades off robustness to gradient noise with convergence rate.

First, consider the Robust Momentum Method. When $\nu = 0$ and there is no gradient noise ($\delta = 0$), the method achieves the fast convergence rate $1 - 1/\sqrt{\kappa}$. Increasing the noise level above $\delta > 0.13$, however, leads to a loss of convergence guarantee. As we increase ν, the convergence rate becomes slower but the method is capable of tolerating larger noise levels. In the limiting case as $\nu = 1 - \frac{1}{2\kappa}$ the Robust Momentum Method becomes the Gradient Method with $\alpha = \frac{1}{L}$ (dashed green line).

It is interesting to note that the Fast Gradient Method has a faster convergence bound than the Robust Momentum Method for noise levels $0.26 < \delta < 0.41$. However, the Fast Gradient Method is also unstable



(a) $\delta = 0$

(b) $\delta = 0.25$

(c) $\delta = 0.5$

Figure 4.4: Simulation of the Robust Momentum Method (RMM) and the Fast Gradient Method (FGM) with relative gradient noise of strength δ and condition ratio $\kappa = 10$. The objective function is the two-dimensional quadratic with gradient (4.18). The measured gradient at each iteration is $u_k = (1-\delta)\nabla f(y_k)$. **(a)** With no noise, all methods are stable and the RMM with $\nu = 0$ is the fastest. **(b)** With more noise, the RMM with $\nu = 0$, the most fragile possible tuning, is unstable. **(c)** With high noise, only the RMM with $\nu = 0.55$ remains stable. Even FGM is unstable with this much noise.

for $\delta > 0.5$ while the Robust Momentum Method can be tuned so that it converges with noise levels up to $\delta \to 1$.

Numerical Simulations. To illustrate the noise robustness properties of different tunings of the Robust Momentum Method, we compared it to the Fast Gradient Method when applied to a simple two-dimensional quadratic function. We used the gradient

$$\nabla f(y_k) = \begin{bmatrix} m & 0 \\ 0 & L \end{bmatrix} (y_k - x_\star) \tag{4.18}$$

where the gradient noise is $r_k = -\delta\, \nabla f(y_k)$. See Figure 4.4. The RMM with $\nu = 0$ has the fastest convergence rate in the noiseless case ($\delta = 0$), but quickly diverges when noise is present. The FGM is more robust to noise, but also diverges when the noise magnitude δ is too large. The RMM with $\nu = 0.55$ remains stable for large amounts of noise, although in the absence of noise the convergence rate is slower than both other methods.

Chapter 5

Final remarks and Future Works

This work is focused on stability of feedback interconnected systems. The main result is a necessary and sufficient stability condition for a class of feedback interconnected systems. Our method is constructive, i.e., we provide the procedure to generate the counter example when the conditions of the theorem are violated. This theorem is written for a general semi-inner product spaces and therefore can be specialized into special cases and, as we have shown in Chapter 3, we can recover necessary-and-sufficient results corresponding to well-known stability theorems.

In addition, we used stability results from the literature to design an optimization algorithm we call the Robust momentum method (RMM) for first-order optimization problems. We show that this algorithm can be tuned to trade off worst-case rate of convergence and sensitivity to noise.

We now present some potential directions for future research that could build on the work in this dissertation. The goal is to draw a road map to continue this research. The suggested future works are:

1. Extension of Corollary 3.1 to the case when M and N are allowed to have dynamics.

2. Extension of the work of section 3 to sector-bounded slope-restricted nonlinearities.

3. Model predictive control (MPC) analysis using system analysis tools of Chapter 3.

4. Further improvement and investigations of Robust Momentum Method, i.e, investigation of convergence rate and stability when a different kind of noise is used. For example, additive noise or random noise with bounded variance.

5. Proposing a unifying KYP-style result for equivalence between a frequency-domain inequality (FDI), an inner-product, and a linear matrix inequality (LMI) [64, 77, 78].

6. Extending the work of Chapter 4 to second-order optimization algorithms.

5.1 Extension of Theorem 3.1 to Systems With Dynamic M and N

One way to continue this research is to consider the case when in Corollary 3.1, M and N are functions, i.e., $M = M^* : j\mathbb{R} \to \mathbb{C}^{2\times 2}$. This new result will be a unifying result that can include IQC theorem [13,31,44]. Since there is an isometric isomorphism between $L_2[0, \infty)$ and $\mathcal{H}_2$ [46, p.98], this can be interpreted as the frequency-domain alternative of Corollary 3.1.

5.2 Extension of The Work of Chapter 3 to Sector-Bounded Slope-Restricted Nonlinearities

Another possibility is to continue the work of section 3 to systems for which the nonlinearity is both sector-bounded *and* slope-restricted. Previously Zames and Falb [2] have solved this problem suggesting Zames-Falb multipliers which find *sufficient* conditions for stability. Extending our theorem to this case could provides *necessary and sufficient conditions* for input-output stability of these systems.

Note that the family of functions that hold in a sector-bounded slope-restricted constraint is a subset of the family of functions that hold in a just sector-bounded constraint with the same bounds. This means that although the same conditions of the "just sector-bounded" case will still guarantee stability (since every sector-bounded slope-restricted system is also a sector-bounded system), but the conditions are too conservative and finding a necessary and sufficient condition for stability of this class of systems means that we need to eliminate this conservativeness.

Some good references to work on this problem are: [5, Thm. VI.5.30], [79, Thm. 4.3], [2,4,80–82,82–85]. Note that a lot of these references are shared with multiplier's theory [5, §VI.9]. Multipliers' theory, was originally proposed by O'shea [8, 9] and later formalized in the form of Zames-Falb multipliers [2], use a (possibly non-causal) LTI system to reduce conservatism. See [4] for a good literature review on the theory of multipliers.

5.3 Model Predictive Control (MPC) Analysis Using System Analysis Tools

Robustness analysis of MPC problems can be assessed using the results of Chapter 3 [86, 87]. There has been few works in the literature that focus on model predictive controllers which are subject to dynamic uncertainties, represented using an IQC. Other good resources for works that use a framework similar to this dissertation to assess robust stability of MPC problems are [88–93].

5.4 Further Investigations of Robust Momentum Method (RMM)

In Chapter 4, the proposed algorithm is only investigated for a certain kind of noise, i.e., multiplicative noise. A possible idea is to observe the behavior of this algorithm when is exposed to other kinds of noise.

Article	**FDI**	**Implication**	**LMI**	**Multiplier**	**Stable**	$\mathscr{C}$	$\mathscr{O}$	**eig.**	**Notes**
Willems [77]	s	$\Longleftarrow$	p.s.d	st	N	Y	N	N	
Willems [77]	ω	$\Longleftrightarrow$	s	st	N	Y	N	N	
Rantzer [64]	ω	$\Longleftrightarrow$	s	st	N	Y	N	Y	
Rantzer [64]	ω	$\Longleftrightarrow$	s	st	N	Y	N	Y	
Rantzer [64]	ω	$\Longleftrightarrow$	p.s.d.	st	Y	Y	N	Y	$M_{11} > 0$
Rantzer [64]	ω	$\Longleftrightarrow$	p.d.	st	Y	Y	Y	Y	$M_{11} > 0$
Wen [94]	ω	$\Longleftrightarrow$	p.d.	$\tilde{I}$	Y	Y	Y	N	SPR
Wen [94]	ω	$\Longleftrightarrow$	p.d.	$\tilde{I}$	Y	Y	Y	N	SPR
Wen [94]	ω	$\Longleftrightarrow$		$\tilde{I}$	Y	Y	Y	N	

Table 5.1: Summary of results in the literature that include a relationship between a frequency-domain inequality, an LMI, and an inner-product. Here $\tilde{I}$ is zero on the main diagonal and identity everywhere else. SPR: Strong positive real. ω means that the frequency domain inequality is written for all $s = j\omega$, and s means that it is written for $s \geq 0$. In LMI column, "s" means symmetric ($P = P^{\mathsf{T}}$), "p.s.d" means positive semi-definite ($P \succeq 0$), and "p.d" means positive definite ($P \succ 0$). "eig." specifies whether or not there are eigenvalues on the $j\omega$ axis. "$\mathscr{C}$" specifies controllability and "$\mathscr{O}$" is observability. "Y" yes, and "N" means no. "st" is static multiplier.

Since we have designed this algorithm using IQCs, it is expected for the algorithm to be robust to noise as any other feedback system which is designed using IQCs.

5.5 Generalized Kalman–Yakubovich–Popov (KYP) Lemma

Another possible future direction for this research is proposing a generalized KYP lemma. This dissertation has heavily used KYP lemma. There are numerous similar results in the literature that connect an LMI, a frequency-domain inequality and sometimes an inner-product. In Table 5.1, a summary of this flavor of results in the literature can be seen. On possible future direction for this work is to obtain a generalized lemma that summarizes important results in the literature.

Remark 5.1. *IQC theorem does not include the case when the nonlinearity belongs to the sector* $[a, b]$ *and* $a > 0$*. The reason is that the Homotopy tool that Megretski and Rantzer [13] are using in their proof requires that the IQC defined by* Π *be satisfied by* $\tau\Delta$ *for all* $\tau \in [0, 1]$*. This means that when we choose* $\tau = 0$*, the IQC should be satisfied by the zero operator* $\Delta = 0$*. Substitute this into the definition of IQC and we will see that we have to have* $\Pi_{11}(j\omega) \geq 0$*. See [65, 95] for more information.*

5.6 Extending the Work of Chapter 4 to Second-Order Optimization Algorithms

In Chapter 4 we showed that first-order optimization algorithms can be written as a robust stability problem. A natural question to ask is about second-order methods, e.g., Newton's method. Note that second-order methods use the Hessian of the function evaluate at the point of interest, in addition to the gradient and therefore Lur'e system with the current definitions cannot describe second-order optimization methods.

Chapter 6

Conclusion

In this dissertation, we considered the problem of robust stability of two systems in feedback. In Chapter 3, we introduced a robust stability result in general semi-inner product spaces. We provided the conditions for which this result is a necessary and sufficient condition for robust stability of systems of the form of equation (3.1). We also developed a procedure that explained how one can construct the counterexample in the case when conditions of the theorem are violated. We showed that this theorem recovers famous classical input-output stability theorems, e.g., conic sector, extended conic sector, small-gain, passivity, etc.

In Chapter 4, we used tools that are available for the problem of robust stability of a feedback system, to design a fast and robust-to-noise first-order optimization algorithm which is being applied to strongly convex functions. We showed that this algorithm on one extreme is as fast as the fastest algorithm available in the literature (Triple momentum algorithm) for this class of problems, and on the other extreme is as robust to (multiplicative) noise as the most robust algorithm in the literature, i.e., Gradient Method with step-size $\alpha = 1/L$, where L is the smoothness parameter of the desired class of m-strongly convex L-smooth functions.

In Chapter 5 we suggested possible future directions for this research. Extension to systems with dynamics multipliers, extension to sector-bounded slope-restricted systems, applying Theorem 3.1 to model predictive control (MPC) problem, and developing a generalized Kalman–Yakubovich–Popov (KYP) lemma are suggested directions to go from this dissertation.

LIST OF REFERENCES

[1] A. Lur'e and V. Postnikov, "On the theory of stability of control systems," *Applied mathematics and mechanics*, vol. 8, no. 3, pp. 246–248, 1944.

[2] G. Zames and P. Falb, "Stability conditions for systems with monotone and slope-restricted nonlinearities," *SIAM Journal on Control*, vol. 6, no. 1, pp. 89–108, 1968.

[3] M. Vidyasagar, *Nonlinear Systems Analysis*, 2nd ed. Society for Industrial and Applied Mathematics, 2002.

[4] J. Carrasco, M. C. Turner, and W. P. Heath, "Zames–Falb multipliers for absolute stability: From O'Shea's contribution to convex searches," *European Journal of Control*, vol. 28, pp. 1–19, 2016.

[5] C. A. Desoer and M. Vidyasagar, *Feedback systems: input-output properties.* Society of Industrial and Applied Mathematics, 2009, vol. 55.

[6] J. C. Willems, "Dissipative dynamical systems— part II: Linear systems with quadratic supply rates," *Archive for rational mechanics and analysis*, vol. 45, no. 5, pp. 352–393, 1972.

[7] ——, "Dissipative dynamical systems— part I: General theory," *Archive for rational mechanics and analysis*, vol. 45, no. 5, pp. 321–351, 1972.

[8] R. O'shea, "A combined frequency-time domain stability criterion for autonomous continuous systems," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 477–484, 1966.

[9] ——, "An improved frequency time domain stability criterion for autonomous continuous systems," *IEEE Transactions on Automatic Control*, vol. 12, no. 6, pp. 725–731, 1967.

[10] R. Brockett, "The status of stability theory for deterministic systems," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 596–606, 1966.

[11] B. Van Scoy, R. A. Freeman, and K. M. Lynch, "The fastest known globally convergent first-order method for minimizing strongly convex functions," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 49–54, Jan 2018.

[12] L. J. Bridgeman and J. R. Forbes, "A comparative study of input–output stability results," *IEEE Transactions on Automatic Control*, vol. 63, no. 2, pp. 463–476, Feb 2018.

[13] A. Megretski and A. Rantzer, "System analysis via integral quadratic constraints," *IEEE Transactions on Automatic Control*, vol. 42, no. 6, pp. 819–830, 1997.

[14] W. M. Haddad, E. G. Collins, and D. S. Bernstein, "Robust stability analysis using the small gain, circle, positivity, and popov theorems: A comparative study," *IEEE Transactions on Control Systems Technology*, vol. 1, no. 4, pp. 290–293, 1993.

[15] B. Anderson, "The small-gain theorem, the passivity theorem and their equivalence," *Journal of the Franklin Institute*, vol. 293, no. 2, pp. 105–115, 1972.

[16] R. Brockett and J. Willems, "Frequency domain stability criteria–part i," *IEEE Transactions on Automatic Control*, vol. 10, no. 3, pp. 255–261, 1965.

[17] E. Jury and B. Lee, "On the stability of a certain class of nonlinear sampled-data systems," *IEEE Transactions on Automatic Control*, vol. 9, no. 1, pp. 51–61, 1964.

[18] ——, "On the absolute stability of nonlinear sample-data systems," *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 551–554, 1964.

[19] Y. Z. Tsypkin, "A criterion for absolute stability of automatic pulse systems with monotonic characteristics of the nonlinear element," in *Soviet Physics Doklady*, vol. 9, 1964, p. 263.

[20] J. Willems and R. Brockett, "Some new rearrangement inequalities having application in stability analysis," *IEEE Transactions on Automatic Control*, vol. 13, no. 5, pp. 539–549, 1968.

[21] R. W. Brockett, "Optimization theory and the converse of the circle criterion." in *National Electronics Conference*, vol. 21, 1965, pp. 697–701.

[22] R. Brockett, "On improving the circle criterion," in *Decision and Control including the 16th Symposium on Adaptive Processes and A Special Symposium on Fuzzy Set Theory and Applications, 1977 IEEE Conference on.* IEEE, 1977, pp. 255–257.

[23] D. J. Hill and P. J. Moylan, "General instability results for interconnected systems," *SIAM Journal on Control and Optimization*, vol. 21, no. 2, pp. 256–279, 1983.

[24] S. Takeda and A. Bergen, "Instability of feedback systems by orthogonal decomposition of $\mathcal{L}_2$," *IEEE Transactions on Automatic Control*, vol. 18, no. 6, pp. 631–636, December 1973.

[25] M. Vidyasagar, "Instability of feedback systems," *IEEE Transactions on Automatic Control*, vol. 22, no. 3, pp. 466–467, Jun 1977.

[26] ——, "$\mathcal{L}_2$-instability criteria for interconnected systems," *SIAM Journal on Control and Optimization*, vol. 15, no. 2, pp. 312–328, 1977.

[27] A. Bergen and S. Takeda, "On instability of feedback systems with a single nonlinear time-varying gain," *IEEE Transactions on Automatic Control*, vol. 16, no. 5, pp. 462–464, 1971.

[28] J. C. Willems, "Stability, instability, invertibility and causality," *SIAM Journal on Control*, vol. 7, no. 4, pp. 645–671, 1969.

[29] M. Fazlyab, A. Ribeiro, M. Morari, and V. M. Preciado, "Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems," *SIAM Journal on Optimization*, vol. 28, no. 3, pp. 2654–2689, 2018.

[30] N. S. Aybat, A. Fallah, M. Gurbuzbalaban, and A. Ozdaglar, "Robust accelerated gradient methods for smooth strongly convex functions," *SIAM Journal on Optimization*, vol. 30, no. 1, pp. 717–751, 2020.

[31] S. Michalowsky, C. Scherer, and C. Ebenbauer, "Robust and structure exploiting optimisation algorithms: an integral quadratic constraint approach," *International Journal of Control*, pp. 1–24, 2020.

[32] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.

[33] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, ser. Applied Optimization. Boston, MA: Kluwer Academic Publishers, 2004, vol. 87.

[34] B. T. Polyak, *Introduction to optimization.* New York: Optimization Software, Publications Division, 1987.

[35] N. Kottenstette, M. J. McCourt, M. Xia, V. Gupta, and P. J. Antsaklis, "On relationships among passivity, positive realness, and dissipativity in linear systems," *Automatica*, vol. 50, no. 4, pp. 1003–1016, 2014.

[36] H. Khalil, *Nonlinear systems.* Upper Saddle River, N.J: Prentice Hall, 2002.

[37] P. Antsaklis, *Linear systems.* Boston, MA: Birkhauser, 2006.

[38] B. D. Anderson, "Internal and external stability of linear time-varying systems," *SIAM Journal on Control and Optimization*, vol. 20, no. 3, pp. 408–413, 1982.

[39] B. Hu and P. Seiler, "Exponential decay rate conditions for uncertain linear systems using integral quadratic constraints," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, pp. 3631–3637, 2016.

[40] R. Boczar, L. Lessard, A. Packard, and B. Recht, "Exponential stability analysis via integral quadratic constraints," *arXiv preprint arXiv:1706.01337*, 2017.

[41] G. Zames, "On the input-output stability of time-varying nonlinear feedback systems part one: Conditions derived using concepts of loop gain, conicity, and positivity," *IEEE Transactions on Automatic Control*, vol. 11, no. 2, pp. 228–238, 1966.

[42] ——, "On the input-output stability of time-varying nonlinear feedback systems–Part II: Conditions involving circles in the frequency plane and sector nonlinearities," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 465–476, 1966.

[43] A. R. Teel, T. Georgiou, L. Praly, and E. Sontag, "Input-output stability," *The control handbook*, vol. 1, pp. 895–908, 1996.

[44] H. Pfifer and P. Seiler, "Integral quadratic constraints for delayed nonlinear and parameter-varying systems," *Automatica*, vol. 56, pp. 36–43, 2015.

[45] S. Z. Khong and A. van der Schaft, "On the converse of the passivity and small-gain theorems for input–output maps," *Automatica*, vol. 97, pp. 58–63, 2018.

[46] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control.* Upper Saddle River, N.J: Prentice Hall, 1996.

[47] S. Z. Khong and C. Y. Kao, "Converse theorems for integral quadratic constraints," *IEEE Trans. Autom. Control*, p. In Press., 2021.

[48] J. S. Shamma, "The necessity of the small-gain theorem for time-varying and nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 36, no. 10, pp. 1138–1147, 1991.

[49] S. Cyrus and L. Lessard, "Unified necessary and sufficient conditions for the robust stability of interconnected sector-bounded systems," in *2019 IEEE 58th Conference on Decision and Control (CDC).* IEEE, 2019, pp. 7690–7695.

[50] A. Megretsky, "Power distribution inequalities in optimization and robustness of uncertain systems," *Journal of Mathematical Systems, Estimation, and Control*, vol. 3, no. 3, pp. 301–319, 1993.

[51] L. J. Bridgeman and J. R. Forbes, "The extended conic sector theorem," *IEEE Transactions on Automatic Control*, vol. 61, no. 7, pp. 1931–1937, 2016.

[52] M. G. Safonov, *Stability and robustness of multivariable feedback systems.* MIT press, 1980.

[53] J. Veenman, C. W. Scherer, and H. Köroğlu, "Robust stability and performance analysis based on integral quadratic constraints," *European Journal of Control*, vol. 31, pp. 1–32, 2016.

[54] J. Carrasco and P. Seiler, "Conditions for the equivalence between IQC and graph separation stability results," *Int. J. Control*, vol. 92, no. 12, pp. 2899–2906, 2019.

[55] M. Fu, S. Dasgupta, and Y. C. Soh, "Integral quadratic constraint approach vs. multiplier approach," *Automatica*, vol. 41, no. 2, pp. 281–287, 2005.

[56] P. Seiler, "Stability analysis with dissipation inequalities and integral quadratic constraints," *IEEE Transactions on Automatic Control*, vol. 60, no. 6, pp. 1704–1709, 2014.

[57] A. Megretski, "Necessary and sufficient conditions of stability: A multiloop generalization of the circle criterion," *IEEE Transactions on Automatic Control*, vol. 38, no. 5, pp. 753–756, 1993.

[58] V. Yakubovich, "S-procedure in nonlinear control theory," *Vestnick Leningrad Univ. Math.*, vol. 4, pp. 73–93, 1997.

[59] M. J. McCourt and P. J. Antsaklis, "Control design for switched systems using passivity indices," in *American Control Conference.* IEEE, 2010, pp. 2499–2504.

[60] J. B. Conway, *A course in functional analysis.* Springer-Verlag, New York, 1990.

[61] M. R. Hestenes, *Optimization theory: The finite dimensional case.* Wiley, 1975.

[62] A. J. van der Schaft, *L_2-Gain and passivity techniques in nonlinear control.* Springer, 2017.

[63] A. R. Teel, "On graphs, conic relations, and input-output stability of nonlinear feedback systems," *IEEE Transactions on Automatic Control*, vol. 41, no. 5, pp. 702–709, 1996.

[64] A. Rantzer, "On the Kalman–Yakubovich–Popov lemma," *Systems & Control Letters*, vol. 28, no. 1, pp. 7–10, 1996.

[65] U. Jönsson, "A lecture on the S-procedure," *Lecture Note at the Royal Institute of technology, Sweden*, vol. 23, pp. 34–36, 2001.

[66] A. Megretski, "How conservative is the circle criterion?" in *Open Problems in Mathematical Systems and Control Theory*, V. Blondel, E. D. Sontag, M. Vidyasagar, and J. C. Willems, Eds. London: Springer London, 1999, pp. 149–151.

[67] A. d'Aspremont, "Smooth optimization with approximate gradient," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1171–1183, 2008.

[68] M. Schmidt, N. L. Roux, and F. R. Bach, "Convergence rates of inexact proximal-gradient methods for convex optimization," in *Advances in neural information processing systems*, 2011, pp. 1458–1466.

[69] O. Devolder, F. Glineur, and Y. Nesterov, "First-order methods of smooth convex optimization with inexact oracle," *Mathematical Programming*, vol. 146, no. 1-2, pp. 37–75, 2014.

[70] ——, "Intermediate gradient methods for smooth convex problems with inexact oracle," CORE Discussion Paper 2013/17, Tech. Rep., 2013.

[71] B. Hu and L. Lessard, "Dissipativity theory for Nesterov's accelerated method," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, 2017, pp. 1549–1557.

[72] B. Hu, P. Seiler, and A. Rantzer, "A unified analysis of stochastic optimization methods using jump system theory and quadratic constraints," in *Proceedings of the 2017 Conference on Learning Theory*, vol. 65, 2017, pp. 1157–1189.

[73] R. Boczar, L. Lessard, and B. Recht, "Exponential convergence bounds using integral quadratic constraints," in *IEEE Conference on Decision and Control (CDC)*, 2015, pp. 7516–7521.

[74] J. Bao and P. L. Lee, *Process control: the passive systems approach.* London: Springer-Verlag, 2007.

[75] E. de Klerk, F. Glineur, and A. B. Taylor, "On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions," *Optimization Letters*, vol. 11, no. 7, pp. 1185–1199, 2017.

[76] ——, "Worst-case convergence analysis of gradient and Newton methods through semidefinite programming performance estimation," *arXiv:1709.05191*, 2017.

[77] J. Willems, "Least squares stationary optimal control and the algebraic Riccati equation," *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 621–634, 1971.

[78] ——, "On the existence of a nonpositive solution to the Riccati equation," *IEEE Transactions on Automatic Control*, vol. 19, no. 5, pp. 592–593, 1974.

[79] J. C. Willems, *The Analysis of feedback systems.* MIT Press, Cambridge, Massachusetts, 1971.

[80] J. Carrasco, W. P. Heath, and A. Lanzon, "Equivalence between classes of multipliers for slope-restricted nonlinearities," *Automatica*, vol. 49, no. 6, pp. 1732–1740, 2013.

[81] M. C. Turner, M. Kerr, and I. Postlethwaite, "On the existence of stable, causal multipliers for systems with slope-restricted nonlinearities," *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2697–2702, 2009.

[82] N. S. Ahmad, J. Carrasco, and W. P. Heath, "A less conservative LMI condition for stability of discrete-time systems with slope-restricted nonlinearities," *IEEE Transactions on Automatic Control*, vol. 60, no. 6, pp. 1692–1697, 2014.

[83] J. Carrasco, W. P. Heath, J. Zhang, N. S. Ahmad, and S. Wang, "Convex searches for discrete-time zames–falb multipliers," *IEEE Transactions on Automatic Control*, vol. 65, no. 11, pp. 4538–4553, 2020.

[84] W. P. Heath, J. Carrasco, and D. A. Altshuller, "Multipliers for nonlinearities with monotone bounds," *arXiv preprint arXiv:2009.09366*, 2020.

[85] X. Xu, B. Acikmese, and M. J. Corless, "Observer-based controllers for incrementally quadratic nonlinear systems with disturbances," *IEEE Transactions on Automatic Control*, 2020.

[86] P. Petsagkourakis, W. P. Heath, J. Carrasco, and C. Theodoropoulos, "Robust stability of barrier-based model predictive control," *IEEE Transactions on Automatic Control*, 2020.

[87] L. Schwenkel, J. Kohler, M. A. Muller, and F. Allgower, "Dynamic uncertainties in model predictive control: Guaranteed stability for constrained linear systems," in *IEEE Conference on Decision and Control (CDC)*, 2020, p. in press.

[88] J. A. Primbs and V. Nevistic, "A framework for robustness analysis of constrained finite receding horizon control," *IEEE Transactions on Automatic Control*, vol. 45, no. 10, pp. 1828–1838, 2000.

[89] P. Petsagkourakis, W. P. Heath, and C. Theodoropoulos, "Stability analysis of piecewise affine systems with multi-model predictive control," *Automatica*, vol. 111, p. 108539, 2020.

[90] W. Heath, G. Li, A. Wills, and B. Lennox, "The robustness of input constrained model predictive control to infinity-norm bound model uncertainty," *IFAC Proceedings Volumes*, vol. 39, no. 9, pp. 495–500, 2006.

[91] W. Heath and A. Wills, "The inherent robustness of constrained linear model predictive control," *IFAC Proceedings Volumes*, vol. 38, no. 1, pp. 71–76, 2005.

[92] P. Petsagkourakis, W. P. Heath, J. Carrasco, and C. Theodoropoulos, "Input-output stability of barrier-based model predictive control," *arXiv preprint arXiv:1903.03154*, 2019.

[93] W. Heath and A. Wills, "Zames-Falb multipliers for quadratic programming," *IEEE Transactions on Automatic Control*, vol. 10, no. 52, pp. 1948–1951, 2007.

[94] J. T. Wen, "Time domain and frequency domain conditions for strict positive realness," *IEEE Transactions on Automatic Control*, vol. 33, no. 10, pp. 988–992, 1988.

[95] P. Seiler, A. Packard, and G. J. Balas, "A dissipation inequality formulation for stability analysis with integral quadratic constraints," in *49th IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 2304–2309.

[96] I. Pólik and T. Terlaky, "A survey of the s-lemma," *SIAM review*, vol. 49, no. 3, pp. 371–418, 2007.

[97] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory.* SIAM, 1994.

Appendix A: S-Lemma proof of Theorem 3.1

In this section we present proof of the sufficiency direction in Theorem 3.1 using S-lemma [50, 61, 96]. Historically, there has been two methods to address the question of necessity of the stability theorems written for system of Figure 1.1:

(i) Proof by contradiction using constructive methods [48] as well as [46, Thm. 9.1], and [3, Lem. 6.6.112].

(ii) Proof using S-lemma [13, 45, 49].

In Chapter 3, a constructive approach was used to prove the necessary direction of the theorem. In this chapter a proof of the necessity direction (Item (2) $\implies$ Item (1)), using S-lemma is provided. Note that using S-lemma has two drawbacks:

1. The proof is not constructive.
2. Using the S-lemma requires the vector space under consideration to be a subspace and this requirement restricts G to be linear.

This being said, we provide the proof for completeness. For a brief history of the S-lemma, see [97, p.23,33] and [65]. Our proof generalizes the version by Hestenes for Hilbert spaces [61, Thm. 7.1, p. 354].

A.1 Using S-Lemma in Theorem 3.1 to Prove (2) $\implies$ (1)

First, let's state the theorem:

Theorem A.1. *Let $\mathcal{V}$ be a semi-inner product space and let $M = M^* \in \mathbb{F}^{2\times 2}$ be indefinite. Suppose $G \in \mathscr{L}(\mathcal{L}_{2e})$ and consider the three following statements.*

1. There exists $N = N^* \in \mathbb{F}^{2\times 2}$ satisfying $M + N \prec 0$ such that G satisfies

$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle \geq 0 \quad \textit{for all } \xi \in \operatorname{dom}(G). \tag{A.1}$$

2. There exists $\gamma > 0$ such that for all (u, y, e), if

$$\left\langle \begin{bmatrix} e_2 \\ y_2 \end{bmatrix}, M \begin{bmatrix} e_2 \\ y_2 \end{bmatrix} \right\rangle \geq 0 \tag{A.2}$$

and (A.3)*,* (A.5)*,* (A.6) *are satisfied, where*

$$e_1 = u_1 + y_2 \tag{A.3}$$

$$y_2 = \Phi e_2 \tag{A.4}$$

$$e_2 = u_2 + y_1 \tag{A.5}$$

$$y_1 = Ge_1, \tag{A.6}$$

then $\|y\| \leq \gamma\|u\|$.

The following equivalences hold: (1) $\iff$ *(2).*

We want to show (1) $\Longleftarrow$ (2). We begin with a generalization of the lossless S-lemma to semi-inner product spaces similar to a Hilbert space version due to Hestenes [61, Thm. 7.1, p. 354] This relies on the following notion of quadratic form.

Definition A.1. *Let* $\mathcal{V}$ *be a real vector space. A* quadratic form Q *is a function* $Q : \mathcal{V} \to \mathbb{R}$ *that has associated with it a function* $\tilde{Q} : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ *such that the following properties hold for all* $x, y, z \in \mathcal{V}$ *and* $a, b \in \mathbb{R}$.

1. $Q(x) = \tilde{Q}(x, x)$
2. $\tilde{Q}(x, y) = \tilde{Q}(y, x)$
3. $\tilde{Q}(x, ay + bz) = a\tilde{Q}(x, y) + b\tilde{Q}(x, z)$
4. $Q(ax + by) = a^2Q(x) + 2ab\tilde{Q}(x, y) + b^2Q(y)$

Lemma A.1. *Let* $\mathcal{V}$ *be a real vector space and let* $S \subseteq \mathcal{V}$ *be a subspace. Let* σ_0 *and* σ_1 *be quadratic forms and suppose there exists* $x^\star \in S$ *such that* $\sigma_1(x^\star) > 0$*. The following statements are equivalent.*

(S1) For all $x \in S$*, we have* $\sigma_1(x) \geq 0 \implies \sigma_0(x) \leq 0$.

(S2) There exists $\tau \geq 0$ *such that for all* $x \in S$*, we have*
$\sigma_0(x) + \tau\sigma_1(x) \leq 0$.

Proof. The proof that (S2) $\Longrightarrow$ (S1) is immediate. To prove the converse, define the sets $\mathcal{N}$ and $\mathcal{W}$, both subsets of $\mathbb{R}^2$:

$$\mathcal{N} := \left\{(u, v) \in \mathbb{R}^2 \;\middle|\; u > 0, v > 0\right\},$$

$$\mathcal{W} := \{(\sigma_0(x), \sigma_1(x)) \mid x \in S\}.$$

The set $\mathcal{W}$ is a convex cone in $\mathbb{R}^2$ [61, Lemma 7.1]. Let (u,v) be an arbitrary element of $\mathcal{W}$. In other words, let $u = \sigma_0(x)$ and $v = \sigma_1(x)$ for some $x \in S$. By assumption, we have $v > 0 \implies v \geq 0 \implies u \leq 0$. It follows that $(u,v) \notin \mathcal{N}$ and so $\mathcal{W} \cap \mathcal{N} = \emptyset$. Since $\mathcal{W}$ and $\mathcal{N}$ are disjoint convex sets and $\mathcal{N}$ is open, there exists a separating hyperplane. In other words, we can find $\lambda \geq 0$ and $\mu \geq 0$, not both zero, such that the half-space $\lambda u + \mu v \leq 0$ contains $\mathcal{W}$. Consequently, $\lambda\sigma_0(x) + \mu\sigma_1(x) \leq 0$ for all $x \in S$. Setting $x = x^\star$, we conclude that $\lambda > 0$. Dividing through by λ and letting $\tau := \frac{\mu}{\lambda}$ completes the proof. ■

We now prove Theorem A.1. We will use Θ to denote a generic tuple $(u,y,e) = (u_1,u_2,y_1,y_2,e_1,e_2) \in \mathcal{X}^6$. Define the sets:

$$\begin{aligned} S'_\Phi &:= \left\{\Theta \in \mathcal{X}^6 \;\middle|\; \text{Equations (3.1a)–(3.1d) hold}\right\}, \\ S &:= \left\{\Theta \in \mathcal{X}^6 \;\middle|\; \text{Equations (3.1a), (3.1c), (3.1d) hold}\right\}. \end{aligned}$$

Note that S'_Φ depends on Φ but S does not. Since G is linear by assumption, it follows that S is a subspace. Moreover, $\bigcup_{\Phi\in\mathcal{C}} S'_\Phi = S$. To see why, first observe that $S'_\Phi \subseteq S$ for all Φ by definition. To prove the opposite inclusion, given any $\Theta \in S$, there exists $\Phi \in \mathcal{C}$ such that $y_2 = \Phi e_2$, which is possible because $\mathcal{C}$ is complete.

Define the quadratic forms on $S \to \mathbb{R}$:

$$\sigma_0(\Theta) := \|y\|^2 - \gamma^2\|u\|^2, \qquad\qquad \sigma_1(\Theta) := \left\langle \begin{bmatrix} e_2 \\ y_2 \end{bmatrix}, M \begin{bmatrix} e_2 \\ y_2 \end{bmatrix} \right\rangle.$$

For any $\Phi \in \mathcal{C}$ and any $e_2 \in \operatorname{dom}(\Phi)$, we can define $\Theta \in S'_\Phi$ using $e_1 = y_1 = 0$, $u_2 = e_2$, and $y_2 = -u_1 = \Phi e_2$. This Θ has the property that $y_2 = \Phi e_2$. Therefore, (3.5) is equivalent to the statement that $\sigma_1(\Theta) \geq 0$ for all $\Theta \in S'_\Phi$.

Item 2 from Theorem A.1 states that for all $\Phi \in \mathcal{C}$ that satisfy $\sigma_1(\Theta) \geq 0$ for all $\Theta \in S'_\Phi$, we have $\sigma_0(\Theta) \leq 0$. But since $\bigcup_{\Phi\in\mathcal{C}} S'_\Phi = S$, Item 2 is equivalent to the statement that for all $\Theta \in S$, we have $\sigma_1(\Theta) \geq 0 \implies \sigma_0(\Theta) \leq 0$.

Since M is indefinite, it must have a positive eigenvalue. So there exists some $v := \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right] \neq 0$ such that $v^\top M v > 0$. Since $\mathcal{X}$ is nontrivial, let ξ_0 be such that $\|\xi_0\| > 0$. Choose $\Theta = (u,y,e) \in S$ using $e_1 = y_1 = 0$, $e_2 = u_2 = a\xi_0$, and $y_2 = -u_1 = b\xi_0$. This choice leads to $\sigma_1(\Theta) = (v^\top M v)\|\xi_0\|^2 > 0$. Since S is a subspace, we may apply Lemma A.1 and conclude that there exists some $\tau \geq 0$ such that

$$\left(\|y\|^2 - \gamma^2\|u\|^2\right) + \tau \left\langle \begin{bmatrix} e_2 \\ y_2 \end{bmatrix}, M \begin{bmatrix} e_2 \\ y_2 \end{bmatrix} \right\rangle \leq 0. \tag{A.7}$$

Now pick $(u,y,e) \in S$ using $u_1 = u_2 = 0$, $y_2 = e_1$, and $e_2 = y_1 = Ge_1$. Using this choice, (A.7) becomes

$$\|Ge_1\|^2 + \|e_1\|^2 + \tau \left\langle \begin{bmatrix} Ge_1 \\ e_1 \end{bmatrix}, M \begin{bmatrix} Ge_1 \\ e_1 \end{bmatrix} \right\rangle \leq 0. \tag{A.8}$$

We must have $\tau > 0$. If instead we had $\tau = 0$, then setting $e_1 = \xi_0$ in (A.8) would lead to an immediate contradiction. Now rearrange (A.8) and obtain

$$\left\langle \begin{bmatrix} Ge_1 \\ e_1 \end{bmatrix}, \left(-\lambda I - M\right) \begin{bmatrix} Ge_1 \\ e_1 \end{bmatrix} \right\rangle \geq 0 \qquad \text{for all } e_1 \in \mathcal{X}.$$

Define $N := -\frac{1}{\tau} I - M$. Then we have $M + N = -\frac{1}{\tau} I \prec 0$, which is (A.1) and so we have proven Item 1 of Theorem A.1. ■

Appendix B: Proving the results in Table 3.2

This chapter provides proofs of the values that are obtained Table 3.2 for M and N. As a reminder, note that M and N are assumed to be two systems in *positive* feedback interconnection (see Figure 3.1). The equations in Corollary 3.1 are:

$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle_T \geq 0, \qquad \text{for all } \xi \in \mathcal{L}_{2\mathrm{e}}, \tag{B.1}$$

and

$$\left\langle \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix}, M \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix} \right\rangle_T \geq 0, \qquad \text{for all } \xi \in \mathcal{L}_{2\mathrm{e}}, \tag{B.2}$$

in addition to the constraint that

$$M + N < 0. \tag{B.3}$$

In the following we will se how based on Corollary 3.1 we can recover values in Table 3.2.

B.1 Small gain theorem

The Table 3.2 suggests

$$M = \begin{bmatrix} \gamma_2 & 0 \\ 0 & -1/\gamma_2 \end{bmatrix}, \quad N = \begin{bmatrix} -1/\gamma_1 & 0 \\ 0 & \gamma_1 \end{bmatrix}$$

with the condition $\gamma_1\gamma_2 < 1$. Substitute these values for M and N in (B.2) and (B.1) and we have:

$$\left\langle \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix}, M \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix} \right\rangle_T \geq 0 \implies \|\Phi\xi\|_T^2 \leq \gamma_2^2 \|\xi\|_T^2,$$

$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle_T \geq 0 \implies \|G\xi\|_T^2 \leq \gamma_1^2 \|\xi\|_T^2,$$

for all $\xi \in \mathcal{L}_{2\mathrm{e}}$, which are describing two systems Φ and G to be gain-bounded, i.e., the condition of the small-gain Theorem [41]. Consequently, the condition (B.3) can be written as

$$M + N = \begin{bmatrix} \gamma_2 - 1/\gamma_1 & 0 \\ 0 & \gamma_1 - 1/\gamma_2 \end{bmatrix} \prec 0$$

and therefore $\gamma_1\gamma_2 < 0$.

B.2 Extended passivity

Substituting values from Table 3.2, into (B.2) and (B.1), we will obtain

$$\left\langle \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix}, M \begin{bmatrix} \xi \\ \Phi\xi \end{bmatrix} \right\rangle_T \geq 0 \implies -\varepsilon_2 \|\xi\|_T^2 + \langle \xi, \, \Phi\xi \rangle_T - \delta_2 \|\Phi\xi\|_T^2 \geq 0,$$

$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle_T \geq 0 \implies -\delta_1 \|G\xi\|_T^2 - \langle G\xi, \, \xi \rangle_T - \varepsilon_1 \|\xi\|_T^2 \geq 0,$$

and the condition $M + N < 0$ means $\varepsilon_2 + \delta_1 \geq 0$ and $\varepsilon_1 + \delta_2 \geq 0$, which is identical to [3, Thm. 6.6.58]. Note that since the feedback interconnection is assumed to be positive in Corollary 3.1, we need to replace G with $-G$ in order to obtain results which use negative feedback convention, e.g. [3, Thm. 6.6.58]. Alternatively, we can replace N from this table with $\tilde{N}$ via multiplying off-diagonal elements by -1 and have:

$$\left\langle \begin{bmatrix} G\xi \\ \xi \end{bmatrix}, N \begin{bmatrix} G\xi \\ \xi \end{bmatrix} \right\rangle_T \geq 0 \implies -\delta_1 \|G\xi\|_T^2 + \langle G\xi, \, \xi \rangle_T - \varepsilon_1 \|\xi\|_T^2 \geq 0$$

$$\implies \langle G\xi, \, \xi \rangle_T \geq \varepsilon_1 \|\xi\|_T^2 + \delta_1 \|G\xi\|_T^2$$

for all $\xi \in \mathcal{L}_{2\mathrm{e}}$.

B.3 Conic sector theorem

For conic sector Theorem, in Table 3.2, we have

$$M = \begin{bmatrix} \frac{-(a+\Delta)(b-\Delta)}{b-a-2\Delta} & \frac{-a-b}{2(b-a-2\Delta)} \\ \frac{-a-b}{2(b-a-2\Delta)} & \frac{-1}{b-a-2\Delta} \end{bmatrix}, \qquad N = \begin{bmatrix} \frac{ab}{b-a+2ab\delta} & \frac{a+b}{2(b-a+2ab\delta)} \\ \frac{a+b}{2(b-a+2ab\delta)} & \frac{(1+a\delta)(1-b\delta)}{b-a+2ab\delta} \end{bmatrix}.$$

Plug in M from above into Corollary 3.1, we obtain

$$\frac{1}{b - a - 2\Delta} \Big[-(a + \Delta)(b - \Delta)\|\xi\|_T^2 - \|\Phi\xi\|_T^2 - (a + b) \langle \xi, \, \Phi\xi \rangle_T \Big] \geq 0. \tag{B.4}$$

To compare these results with classical results in the literature, note that positive sign convention should change to negative feedback convention. One way to do so is to substitute N with $\tilde{N}$ (i.e., multiply off-diagonal elements of N with -1). An alternative is to multiply Φ with -1, and therefore write (B.4) as

$$\frac{1}{b - a - 2\Delta} \Big[-(a + \Delta)(b - \Delta)\|\xi\|_T^2 - \|\Phi\xi\|_T^2 + (a + b) \langle \xi, \, \Phi\xi \rangle_T \Big] \geq 0. \tag{B.5}$$

If $b > a + 2\Delta$ this describes $\Phi \in \mathrm{Cone}[a + \Delta, b - \Delta]$. On the other hand, plugging in the value of N in Corollary 3.1, we obtain

$$\frac{ab}{b - a + 2ab\delta}\|G\xi\|_T^2 + \frac{a + b}{b - a + 2ab\delta} \langle G\xi, \, \xi \rangle_T + \frac{(1 + a\delta)(1 - b\delta)}{b - a + 2ab\delta}\|\xi\|_T^2 \geq 0 \tag{B.6}$$

If $ab > 0$, we have

$$\begin{aligned}
&\frac{1}{b-a+2ab\delta}\Big[\|G\xi\|_T^2 + (\frac{1}{a}+\frac{1}{b})\left\langle G\xi,\, \xi\right\rangle_T + \frac{(1+a\delta)(1-b\delta)}{ab}\|\xi\|_T^2\Big] \\
&= \frac{1}{b-a+2ab\delta}\Big[\|G\xi\|_T^2 - (-\frac{1}{a}-\delta-\frac{1}{b}+\delta)\left\langle G\xi,\, \xi\right\rangle_T + \frac{(1+a\delta)(1-b\delta)}{ab}\|\xi\|_T^2\Big] \\
&= \frac{1}{b-a+2ab\delta}\Big[\|G\xi\|_T^2 - (-\frac{1}{a}-\delta-\frac{1}{b}+\delta)\left\langle G\xi,\, \xi\right\rangle_T + (\frac{1}{a}+\delta)(\frac{1}{b}-\delta)\|\xi\|_T^2\Big] \\
&= \frac{1}{b-a+2ab\delta}\Big[\|G\xi\|_T^2 - (-\frac{1}{a}-\delta-\frac{1}{b}+\delta)\left\langle G\xi,\, \xi\right\rangle_T + (-\frac{1}{a}-\delta)(-\frac{1}{b}+\delta)\|\xi\|_T^2\Big] \geq 0.
\end{aligned}$$

Note that and exterior conic system in $[\alpha, \beta]$ can be described with

$$\|G\xi\|_T^2 - (\alpha+\beta)\left\langle G\xi,\, \xi\right\rangle_T + \alpha\beta\|\xi\|_T^2 \geq 0.$$

Therefore $G \in \text{Excone}[\alpha, \beta]$ with $\alpha = -\frac{1}{a} - \delta$ and $\beta = -\frac{1}{b} + \delta$. If $a = 0$ and $b > 0$, from (B.6) we get $\left\langle G\xi,\, \xi\right\rangle_T + (\frac{1}{b} - \delta)\|\xi\|_T^2 \geq 0$ and therefore $G \in \text{Excone}[\alpha, \beta]$, with $\alpha = -\frac{1}{a} - \delta$ and $\beta = -\frac{1}{b} + \delta$.

If $ab < 0$ we can write

$$\begin{aligned}
&\frac{1}{b-a+2ab\delta}\Big[-\|G\xi\|_T^2 - (\frac{1}{a}+\frac{1}{b})\left\langle G\xi,\, \xi\right\rangle_T - \frac{(1+a\delta)(1-b\delta)}{ab}\|\xi\|_T^2\Big] \\
&\frac{1}{b-a+2ab\delta}\Big[-\|G\xi\|_T^2 + (-\frac{1}{a}-\delta-\frac{1}{b}+\delta)\left\langle G\xi,\, \xi\right\rangle_T - (\frac{1}{a}+\delta)(\frac{1}{b}-\delta)\|\xi\|_T^2\Big] \\
&\frac{1}{b-a+2ab\delta}\Big[-\|G\xi\|_T^2 + (-\frac{1}{a}-\delta-\frac{1}{b}+\delta)\left\langle G\xi,\, \xi\right\rangle_T - (-\frac{1}{a}-\delta)(-\frac{1}{b}+\delta)\|\xi\|_T^2\Big] \geq 0
\end{aligned}$$

Remember that a conic sector system in $[\alpha, \beta]$ can be described as

$$-\|G\xi\|_T^2 + (\alpha+\beta)\left\langle G\xi,\, \xi\right\rangle_T - \alpha\beta\|\xi\|_T^2 \geq 0$$

Therefore $G \in \text{Cone}[\alpha, \beta]$ with $\alpha = -\frac{1}{b} + \delta$ and $\beta = -\frac{1}{a} - \delta$. If $a < 0$ and $b = 0$, from (B.6) we get $-\left\langle G\xi,\, \xi\right\rangle_T - (\frac{1}{a} + \delta)\|\xi\|_T^2 \geq 0$ which means $G \in \text{Cone}[-\infty, -\frac{1}{a} - \delta]$. The condition $M + N \prec 0$ can then be written as

$$M+N = \begin{bmatrix} \frac{-(a+\Delta)(b-\Delta)}{b-a-2\Delta} & \frac{-a-b}{2(b-a-2\Delta)} \\ \frac{-a-b}{2(b-a-2\Delta)} & \frac{-1}{b-a-2\Delta} \end{bmatrix} + \begin{bmatrix} \frac{ab}{b-a+2ab\delta} & \frac{a+b}{2(b-a+2ab\delta)} \\ \frac{a+b}{2(b-a+2ab\delta)} & \frac{(1+a\delta)(1-b\delta)}{b-a+2ab\delta} \end{bmatrix}.$$

If $\delta = 0$ and $\Delta > 0$,

$$M+N = \begin{bmatrix} \frac{-(a+\Delta)(b-\Delta)}{b-a-2\Delta} & \frac{-a-b}{2(b-a-2\Delta)} \\ \frac{-a-b}{2(b-a-2\Delta)} & \frac{-1}{b-a-2\Delta} \end{bmatrix} + \begin{bmatrix} \frac{ab}{b-a} & \frac{a+b}{2(b-a)} \\ \frac{a+b}{2(b-a)} & \frac{1}{b-a} \end{bmatrix}.$$

Note that

$$\det(M+N) = \frac{\Delta^2}{(a-b)(a-b+2\Delta)},$$

which is positive since $b > a$. The first element on the main diagonal is $\frac{-\Delta(a^2+b(b-\Delta)+a\Delta)}{(a-b)(a-b+2\Delta)}$ which is negative, the second element on the main diagonal is $\frac{-2\Delta}{(b-a)(b-a-2\Delta)}$ which is negative since $b > a$. Therefore $M+N \prec 0$.

If $\Delta = 0$ and $\delta > 0$,

$$M + N = \begin{bmatrix} \frac{-ab}{b-a} & \frac{-a-b}{2(b-a)} \\ \frac{-a-b}{2(b-a)} & \frac{-1}{b-a} \end{bmatrix} + \begin{bmatrix} \frac{ab}{b-a+2ab\delta} & \frac{a+b}{2(b-a+2ab\delta)} \\ \frac{a+b}{2(b-a+2ab\delta)} & \frac{(1+a\delta)(1-b\delta)}{b-a+2ab\delta} \end{bmatrix}$$

The first element of the main diagonal is $\frac{-2a^2b^2\delta}{(b-a)(b-a+2ab\delta)}$ which is negative and the second element on the main diagonal is

$$\frac{\delta(-a^2 - b^2 + a^2b\delta - ab^2\delta)}{(b-a)(b-a+2ab\delta)},$$

which is negative when $b > a$. In addition

$$\det(M+N) = \frac{a^2b^2\delta^2}{(b-a)(b-a+2ab\delta)},$$

which is positive when $b > a$ and hence $M + N \prec 0$ and we recover the Conic sector Theorem [41].

B.4 Exterior conic theorem

From Table 3.2, we have

$$N = \begin{bmatrix} \frac{-ab}{b-a-2ab\delta} & \frac{-a-b}{2(b-a-2ab\delta)} \\ \frac{-a-b}{2(b-a-2ab\delta)} & \frac{-(1-a\delta)(1+b\delta)}{b-a-2ab\delta} \end{bmatrix}, \qquad M = \begin{bmatrix} \frac{(a-\Delta)(b+\Delta)}{b-a+2\Delta} & \frac{a+b}{2(b-a+2\Delta)} \\ \frac{a+b}{2(b-a+2\Delta)} & \frac{1}{b-a+2\Delta} \end{bmatrix}.$$

Using this value of M in Corollary 3.1 results

$$\frac{1}{b-a+2\Delta}\Big[(a-\Delta)(a+\Delta)\|\xi\|_T^2 + \|\Phi\xi\|_T^2 + (a+b)\,\langle \xi,\, \Phi\xi\rangle_T\Big] \geq 0.$$

To compare this equation with results in the literature that assume negative feedback, we need to replace Φ with $-\Phi$ and therefore

$$\frac{1}{b-a+2\Delta}\Big[(a-\Delta)(a+\Delta)\|\xi\|_T^2 + \|\Phi\xi\|_T^2 - (a+b)\,\langle \xi,\, \Phi\xi\rangle_T\Big] \geq 0,$$

which means $\Phi \in \text{Excone}[\alpha, \beta]$ with $\alpha = a - \Delta$ and $\beta = b + \Delta$. Using N in Corollary 3.1 results

$$\frac{1}{b-a-2ab\delta}\Big[-ab\|G\xi\|_T^2 - (1-a\delta)(1+b\delta)\|\xi\|_T^2 - (a+b)\,\langle \xi,\, G\xi\rangle_T\Big] \geq 0. \tag{B.7}$$

If $ab > 0$ and assuming $b > a$

$$\frac{1}{b-a-2ab\delta}\Big[-\|G\xi\|_T^2 - (-\frac{1}{a}+\delta)(-\frac{1}{b}-\delta)\|\xi\|_T^2 + (-\frac{1}{a}-\frac{1}{b})\,\langle \xi,\, G\xi\rangle_T\Big] \geq 0,$$

and therefore $G \in \text{Cone}[-\frac{1}{a}+\delta, -\frac{1}{b}-\delta]$. If $a < 0$ and $b = 0$, from (B.7) we get

$$(\frac{1}{a}-\delta)\|\xi\|_T^2 + \langle \xi,\, G\xi\rangle_T \geq 0$$

which describes $G \in \text{Cone}[-\frac{1}{a}+\delta, -\infty]$. If $ab < 0$ and assuming $b > a$ we have

$$\frac{1}{b-a-2ab\delta}\Big[\|G\xi\|_T^2 + (-\frac{1}{a}+\delta)(-\frac{1}{b}-\delta)\|\xi\|_T^2 - (-\frac{1}{a}-\frac{1}{b})\,\langle \xi,\, G\xi\rangle_T\Big] \geq 0,$$

and therefore $G \in \text{Excone}[-\frac{1}{b} - \delta, -\frac{1}{a} + \delta]$. If $a = 0$ and $b > 0$, from (B.7) we get

$$-(\frac{1}{b} + \delta)\|\xi\|_T^2 - \langle \xi,\, G\xi \rangle_T \geq 0,$$

and therefore $G \in \text{Excone}[-\frac{1}{b} - \delta, -\infty]$. To check the condition $M + N \prec 0$, note that

$$M + N = \begin{bmatrix} \frac{-ab}{b-a-2ab\delta} & \frac{-a-b}{2(b-a-2ab\delta)} \\ \frac{-a-b}{2(b-a-2ab\delta)} & \frac{-(1-a\delta)(1+b\delta)}{b-a-2ab\delta} \end{bmatrix} + \begin{bmatrix} \frac{(a-\Delta)(b+\Delta)}{b-a+2\Delta} & \frac{a+b}{2(b-a+2\Delta)} \\ \frac{a+b}{2(b-a+2\Delta)} & \frac{1}{b-a+2\Delta} \end{bmatrix}.$$

If $\delta = 0$ and $\Delta > 0$, we have $\det(M + N) = \frac{\Delta^2}{(b-a)(b-a+2\Delta)}$ which is positive since $b > a$. On the other hand

$$(M + N)_{22} = \frac{-2\Delta}{(b-a)(b-a+2\Delta)}, \qquad \text{and} \qquad (M + N)_{11} = \frac{-\Delta(a^2 + b^2 - a\Delta + b\Delta)}{(b-a)(b-a+2\Delta)}$$

which are both negative and therefore $M+N \prec 0$. If $\Delta = 0$ and $\delta > 0$, we have $\det(M+N) = \frac{a^2b^2\delta^2}{(b-a)(b-a-2ab\delta)}$ which is positive, and

$$(M + N)_{22} = \frac{-\delta(a^2 + b^2 + a^2b\delta - ab^2\delta)}{(b-a)(b-a-2ab\delta)}, \qquad \text{and} \qquad (M + N)_{11} = \frac{-2a^2b^2\delta}{(b-a)(b-a-2ab\delta)}$$

which are both negative and therefore $M + N \prec 0$. Therefore we've recovered the extended conic sector theorem [12].