Input-Output Stability of First-Order Optimization Algorithms: A Passivity-Based Gain-Scheduling Approach

by

Sepehr Moalemi

Department of Mechanical Engineering
McGill University, Montreal

August 2025

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Table of Contents

Abstract

The focus of this thesis is a control-theoretic analysis of first-order optimization algorithms, such as gradient descent (GD) and its accelerated variants. The control interpretation of such first-order algorithms consists of casting them as a Lur'e problem, where the algorithm is represented as the interconnection of a linear controller in negative feedback with the gradient. Consequently, standard input-output theory, in particular, the passivity theorem, can be used to analyze the input-output stability of the interconnection. This passivity-based approach is particularly useful, as it is robust to model uncertainties, such as gradient uncertainty that does not violate passivity. Moreover, the stability of optimization algorithms with varying hyperparameters can be analyzed through the lens of passivity-based gain-scheduling techniques. The novel contributions of this thesis come in two parts.

First, it is shown that GD, Polyak's heavy ball (HB), Nesterov's accelerated gradient (NAG), and triple momentum (TM) methods can each be represented as a passive controller in negative feedback with the gradient to be minimized. As such, for a class of functions with sector-bounded gradient, the passivity theorem can be used to guarantee the input-output stability as well as the global convergence of these algorithms. Furthermore, compared with existing results, this approach provides a tighter upper bound on the largest allowable step size by guaranteeing the input-output stability of the GD controller.

Second, this thesis introduces the gain-scheduling of discrete-time quadratic supply rate (QSR)-dissipative subsystems. The scheduling functions considered here take the form of matrices, generalizing the scalar scheduling functions used in the literature. Furthermore, this thesis extends existing gain-scheduling results to include a broader class of QSR-dissipative systems. Notably, given that passivity is a special case of QSR-dissipativity, it is shown that the matrix-scheduling of passive subsystems results in an overall passive system. As such, the proposed gain-scheduling architecture, in tandem with the passive GD, HB, NAG, and TM controllers, can be used to represent varying hyperparameters and propose new variations of these algorithms.

Résumé

Cette thèse examine la stabilité des algorithmes d'optimisation de premier ordre, tels que la descente de gradient (GD) et ses variantes accélérées, à l'aide de la théorie du contrôle. L'interprétation de ces algorithmes dans le cadre de la théorie du contrôle consiste à les reformuler comme un problème de Lur'e, où l'algorithme est représenté comme l'interconnexion d'un contrôleur linéaire en rétroaction négative avec le gradient. Par conséquent, la théorie classique des systèmes entrée-sortie, en particulier le théorème de passivité, peut être utilisée pour analyser la stabilité entrée-sortie de cette interconnexion. Cette approche fondée sur la passivité est particulièrement utile, car elle est robuste aux incertitudes du modèle, telles qu'une incertitude sur le gradient qui ne viole pas la propriété de passivité. De plus, la stabilité des algorithmes d'optimisation à hyperparamètres variables peut être analysée sous l'angle de techniques de commande à gains programmés fondées sur la passivité. Les contributions originales de cette thèse s'articulent en deux volets.

Premièrement, il est démontré que GD, la méthode *heavy ball* de Polyak (HB), la méthode gradient accéléré de Nesterov (NAG) et la méthode à *triple momentum* (TM) peuvent chacune être représentées comme des contrôleurs passifs en rétroaction négative avec le gradient de la fonction à minimiser. Ainsi, pour une classe de fonctions dont le gradient est borné dans un secteur, le théorème de passivité peut être utilisé pour garantir la stabilité entrée-sortie ainsi que la convergence globale de ces algorithmes. De plus, comparativement aux résultats existants, cette approche offre une borne supérieure plus serrée pour la taille de pas maximale autorisée, tout en garantissant la stabilité entrée-sortie du contrôleur de descente de gradient.

Deuxièmement, cette thèse introduit la commande à gains programmés pour des sous-systèmes dissipatifs en temps discret de type QSR (*Quadratic Supply Rate*). Les fonctions d'ordonnancement considérées ici prennent la forme de matrices, généralisant ainsi les fonctions d'ordonnancement scalaires utilisées dans la littérature. En outre, cette thèse étend les résultats existants de la commande à gains programmés, englobant ainsi une classe plus large de systèmes dissipatifs de type QSR. La passivité étant un cas particulier de la dissipativité QSR, il est démontré que l'ordonnancement matriciel de sous-systèmes passifs aboutit à un système passif dans son ensemble. Ainsi, l'architecture de commande à gains programmés proposée, combinée aux contrôleurs passifs GD, HB, NAG et TM, permet d'intégrer des hyperparamètres variables et de proposer de nouvelles variantes de ces algorithmes.

Acknowledgements

First, I would like to thank my advisor, Professor James Richard Forbes, whose passion for all things math related has been a constant source of inspiration. Throughout my time at McGill, Prof. Forbes has gone above and beyond as an instructor, advisor, and mentor to make sure I have the tools and knowledge to succeed. Little did I know that "buying into" passivity theory would alter my academic trajectory to such a degree. It has been said that we stand on the shoulders of giants; in my case, those shoulders have been taller than most.

Second, I would like to thank every member of the DECAR team for creating such a welcoming social and academic environment. It was this environment that encouraged me to come to the lab almost every day.

Finally, I would like to thank my family who have always supported me in my academic pursuits. I would not be where I am today without such role models in my life. And of course, I would like to thank my partner, Emma N. Friesen, for pushing me to be the best version of myself at Pickleball (and in life).

Preface

The novel contributions of this thesis are listed in this preface where any contribution that has appeared in a publication is referenced.

- Chapter 4
  - Introducing the passivity-based analysis of the GD algorithm for a class of functions with sector-bounded gradients and a unique global minimizer, $f \in \mathcal{S}_{m,L}$ with $0 < m \leq L$ [1].
  - Relaxing the assumptions in [1] to $m = 0$ and using the weak passivity theorem to guarantee the global convergence of the GD algorithm.
- Chapter 5
  - Guaranteeing the global convergence of the HB, NAG, and TM methods for $f \in \mathcal{S}_{0,L}$ with $0 < L$ using the weak passivity theorem.
- Chapter 6
  - Generalizing the gain-scheduling architecture presented in [2] by introducing scheduling matrices and formulating the discrete-time counterpart of the results. Note, the presented matrix-scheduling architecture is the discrete-time counterpart of the author's work in [3, 4].
  - GD with varying step size as a gain-scheduled system using the proposed matrix-scheduling architecture.
  - Introducing a variation of the GD method with variable step size, using a passivity-based gain-scheduled control architecture [1].
  - Gain-scheduling passive first-order optimization algorithms.

All proofs, plots, and tables were produced by the author under the supervision of Prof. James Richard Forbes. The sole exception is Lemma A.1, where Prof. Xiao-Wen Chang provided helpful insights to generalize the proof.

List of Figures

Figure

List of Tables

Table

List of Appendices

Appendix

List of Abbreviations

List of Symbols

$\mathbb{Z}$	the set of integers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	the set of real n-dimensional column matrices
$\mathbb{R}^{m\times n}$	the set of real $m \times n$ matrices
$\mathbb{R}_{>0}$, $\mathbb{R}_{\geq 0}$, ...	the set of positive real numbers, non-negative real numbers, etc.
$\mathbb{S}^n$	the set of real symmetric $n \times n$ matrices
$\mathcal{N}$	a set of indices
x	a scalar
$\mathbf{x}$	a column matrix
$\mathbf{X}$	a matrix
$\mathbf{1}$	the identity matrix
$\mathbf{0}$	a zero matrix
$(\cdot)^{\mathsf{T}}$	the transpose operator
$(\cdot)^{-1}$	the inverse operator
$\mathrm{diag}(\cdot)$	a block diagonal matrix of its arguments
$\det(\cdot)$	the determinant of a matrix
$\mathrm{rank}(\cdot)$	the rank of a matrix
$\sigma_{\max}(\cdot)$	the maximum singular value of a matrix
$\lambda_{\max}(\cdot)$	the maximum eigenvalue of a matrix
$*$	the symmetric part of a matrix
$\otimes$	the Kronecker product
$\lvert\cdot\rvert$	an absolute value
$\langle\cdot,\cdot\rangle$	an inner product
$\langle\cdot,\cdot\rangle_T$	a truncated inner product
$\lVert\cdot\rVert$	a norm
$\lVert\cdot\rVert_T$	a truncated norm

$\nabla(\cdot)$	the gradient operator
$\mathbf{A} \succ 0$, $\mathbf{A} \succeq 0$, ...	the matrix $\mathbf{A}$ is positive definite, positive semidefinite, etc.
$\mathbb{S}^n_+$, $\mathbb{S}^n_{++}$, ...	the set of real symmetric positive semidefinite, positive definite, etc. matrices
ℓ_2	the space of square-summable signals
ℓ_{2e}	the extended space of square-summable signals
ℓ_∞	the space of bounded signals
$\mathcal{X} \to \mathcal{Y}$	a mapping from $\mathcal{X}$ to $\mathcal{Y}$
$\boldsymbol{\mathcal{G}} : (\cdot) \to (\cdot)$	an operator
$\mathbf{G}(z)$	a transfer matrix
$\mathcal{F}_{m,L}$	the set of continuously differentiable functions that are m-strongly convex and L-smooth
$\mathcal{S}_{m,L}$	the set of continuously differentiable functions with unique a global minimizer and a sector-bounded gradient

Part 0

Introduction

Chapter 1

Introduction

1.1 Motivation

Optimization is an indispensable tool used in many branches of engineering. The growing complexity of our modern systems and data-driven applications demands optimization algorithms that are robust to model uncertainties, noise, and gradient estimation errors. The inherent accuracy-robustness trade-off of these algorithms motivates the need for a systematic approach to analyze them. By leveraging powerful tools from robust control theory, such as passivity and gain-scheduling, this thesis provides a systematic approach to analyze the stability of first-order optimization algorithms and proposes new variations of these algorithms. To this end, Part 1 of this thesis presents a novel passivity-based analysis of first-order optimization algorithms, while Part 2 generalizes existing gain-scheduling architectures and applies them to the case of gain-scheduling of first-order optimization algorithms.

1.2 Background and Related Work

1.2.1 Optimization as a Control Problem

Input-output stability theorems, such as the passivity, small gain, and conic sector theorems, have been widely used to analyze and guarantee the $\mathcal{L}_2$-stability of feedback interconnections. The passivity theorem ensures closed-loop $\mathcal{L}_2$-stability of a passive system connected in a negative feedback interconnection with a very strictly passive (VSP) system [5]. The systems within the feedback loop are permitted to be time-varying or nonlinear. Moreover, precise knowledge of the system parameters is not required so long as the systems remain passive and VSP, respectively, in the face of model uncertainty.

Recently, popular first-order optimization methods such as gradient descent, Polyak's heavy-ball [6], and Nesterov's accelerated gradient method [7], have been analyzed through the lens of feedback control in a discrete-time setting. In particular, in [8], for a function with sector-bounded gradient, the circle criterion is used to derive expressions for the parameters of the heavy ball method that guarantee its global convergence. This result was extended in [9] by incorporating an additional parameter, leading to the new generalized accelerated gradient method. The small gain theorem is used in [10] to relate the input-output gain computations of a defined complementary sensitivity function to the convergence properties of gradient descent methods. Additionally, [10] demonstrates a connection between the step size selection of the gradient method and $\mathcal{H}_\infty$ state feedback synthesis. For a class of convex constraint-coupled optimization problems, a passivity-based analysis of the alternating direction method of multipliers (ADMM) is provided in [11]. Dissipativity theory and Lyapunov-based approaches for algorithm analysis have also been studied in [12, 13]. Other robust control tools such as integral quadratic constraints (IQCs) have been used in the analysis and synthesis of novel first-order iterative optimization methods. Within the context of optimization, the IQC framework is first used in [14] to derive numerical upper bounds on convergence rates of first-order methods. For the class of m-strongly convex and L-smooth functions, a graphical design procedure based on IQCs is used to derive the optimal parameters for the triple momentum method introduced in [15]. The aforementioned works have mostly been limited to the case of first-order methods with fixed hyperparameters. Notably, [16] appears to be the only widely recognized work that considers varying step sizes, using the IQC framework to model the gradient descent method as a linear parameter-varying system.

1.2.2 Passivity-Based Gain-Scheduled Control

When controlling a nonlinear system, it is possible to design linear controllers using the linearized model of a system about an operating point. However, the linearized model of the plant may not adequately capture the behavior of the nonlinear system over a wide range of operating conditions. Therefore, a single linear controller designed using a single linearized plant may not realize adequate closed-loop performance, or even closed-loop stability, across a wide range of operating conditions. At a high level, gain-scheduling is a “divide and conquer” approach to control design, where a set of linear subcontrollers are designed, each about a different linearization point across the operating range of the nonlinear system. The linear subcontrollers are then interpolated or scheduled to form a gain-scheduled controller that realizes acceptable performance, ideally with closed-loop stability guarantees as well.

The stability of gain-scheduled controllers has been studied using passivity, conicity, and dissipativity theory. Within the context of passivity-based control, a useful passivity-preserving gain-scheduling architecture was introduced in [2]. Using this architecture, the gain-scheduling of strictly positive real (SPR) subcontrollers is shown to be input strictly passive (ISP) [2] and VSP [17]. An alternative passivity-based gain-scheduling architecture accounting for actuator saturation is presented in [18]. Gain-scheduling VSP controllers with affine dependence on plant parameters is presented in [19]. Gain-scheduling conic systems, and relying on the conic sector theorem to ensure $\mathcal{L}_2$-stability of the closed-loop system, is considered in [20]. In [21], quadratic supply rate (QSR)-dissipative properties of non-square QSR-dissipative systems are shown to be preserved under the same gain-scheduling architecture of [2, 17].

The gain-scheduling techniques in [2, 17–22] all consider scalar scheduling signals. In [23], the notion of extended passive systems was shown using a row of scalar scheduling signals. Scalar scheduling signals affect the entire input-output map of the subcontrollers. When controlling a multiple-input multiple-output (MIMO) system, each control variable may require different gain scheduling to ensure acceptable performance. Additionally, it might be natural to gain-schedule based on two or more independent exogenous variables. This motivates the use of scheduling matrices that effectively introduce more scheduling parameters to allow for additional flexibility in the scheduling of the subcontrollers. Moreover, the gain-scheduling results presented in [2], along with its subsequent extensions, focus exclusively on continuous-time systems. While extending these results to discrete-time systems is straightforward, this has not been explicitly addressed in the literature.

1.3 Objectives

This thesis has three main objectives. First, to develop a passivity-based analysis of first-order optimization algorithms that decouples the analysis of the algorithm from the properties of the objective function. This decoupling allows portions of the analysis to be reused when studying new algorithms or objective functions, without the need to repeat the entire procedure. Second, to generalize the gain-scheduling architecture presented in [2] by introducing scheduling matrices and formulating the discrete-time counterpart of the results. Finally, to leverage the proposed discrete-time gain-scheduling architecture and analyze the stability and convergence properties of first-order optimization algorithms with varying hyperparameters using a passivity-based approach.

1.4 Outline and Contributions

General concepts, definitions, and theorems used throughout this thesis, including those related to linear algebra, signals, and input-output stability, are presented in Chapter 2. Chapter specific preliminaries are presented at the beginning of each chapter. The remainder of this thesis is organized as follows.

Part 1 presents a novel passivity-based analysis of first-order optimization algorithms. Firstly, Chapter 3 formally defines the concept of passivity and the passivity theorem. The control interpretation of the gradient descent algorithm is then reviewed in Chapter 4. The contributions of this chapter is derived from [1], where the passivity-based analysis of the gradient descent algorithm is presented. The main result of this chapter uses the weak passivity theorem to provide a tighter upper bound on the largest allowable step size by guaranteeing the input-output stability of the gradient descent controller. The extension of the passivity-based analysis to momentum-based methods is presented in Chapter 5. The highlight of this chapter is to guarantee the input-output stability, as well as the global convergence, of popular momentum-based methods for more general objective functions than those considered in the control literature.

Part 2 presents the gain-scheduling of QSR-dissipative systems. The contributions of Chapter 6 are the discrete-time counterparts of those in [3, 4], where a novel matrix-gain-scheduling architecture was introduced. The main results of this chapter are the applications of gain-scheduling of first-order optimization algorithms. To this end, this chapter presents a control interpretation of the gradient descent algorithm with varying step size as a gain-scheduled system. Moreover, a new variation of the gradient descent algorithm is proposed, where both the input and output of the gradient are scaled. Finally, the gain-scheduling of passive first-order optimization algorithms is discussed, in which the scheduling signals are used to interpolate or switch between different first-order optimization algorithms.

Chapter 7 concludes this thesis and discusses potential future research directions. For a summary of the contributions of this thesis along with appearances in publications, see the Preface.

Chapter 2

Mathematical Preliminaries

The two main themes of this thesis are the use of norm inequalities and input-output stability. Section 2.1 presents the necessary linear algebra preliminaries, including matrix norms and inequalities, which are used throughout this thesis. Section 2.2 reviews the definitions of signals and systems, as well as input-output stability, which are essential for the passivity-based analysis proposed in the subsequent chapters.

2.1 Linear Algebra

This section presents the necessary linear algebra preliminaries, including matrix norms, useful inequalities, and definiteness of matrices, which are used throughout this thesis.

2.1.1 Vector Norms and Matrix Norms

Definition 2.1 (Vector Norm [24])**.** Let $\mathcal{X}$ be a vector space. A real-valued function $\|\cdot\|$ defined on $\mathcal{X}$ is said to be a *norm* on $\mathcal{X}$ if it satisfies the following properties:

(i) $\|\mathbf{x}\| \geq 0$ (positivity);

(ii) $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$ (positive definiteness);

(iii) $\|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|$, for any scalar α (homogeneity);

(iv) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (triangle inequality).

Definition 2.2 (Vector p-Norm)**.** For $\mathbf{x} \in \mathbb{R}^n$, the vector p-norm of $\mathbf{x}$ is defined as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \qquad 1 \leq p < \infty.$$

In particular, for $p = 2$, the vector 2-norm is defined as

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} = \sqrt{\mathbf{x}^\mathsf{T}\mathbf{x}} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle},$$

and for $p = \infty$, the vector ∞-norm is defined as

$$\|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i|.$$

Definition 2.3 (Induced Matrix p-Norm [24])**.** For $\mathbf{A} \in \mathbb{R}^{m \times n}$, the matrix norm induced by a vector p-norm is defined as $\|\mathbf{A}\|_p = \sup_{\mathbf{x} \neq \mathbf{0}} \|\mathbf{A}\mathbf{x}\|_p / \|\mathbf{x}\|_p$. The special case of $p = 2$ can be written as $\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^\mathsf{T}\mathbf{A})} = \sigma_{\max}(\mathbf{A})$.

2.1.2 Useful Inequalities

Lemma 2.1 (Cauchy–Schwarz Inequality [24])**.** Let $\mathcal{X}$ be an inner product space and let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. Then,

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \le \|\mathbf{x}\|_2 \|\mathbf{y}\|_2.$$

Moreover, $|\langle \mathbf{x}, \mathbf{y} \rangle| = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$ if and only if $\mathbf{x} = \alpha \mathbf{y}$ for some $\alpha \in \mathbb{R}_{\ge 0}$.

Lemma 2.2 (Rayleigh Inequality [25])**.** Given a symmetric matrix $\mathbf{Q} \in \mathbb{S}^n$, it follows that

$$\lambda_{\min}(\mathbf{Q}) \|\mathbf{x}\|_2^2 \le \mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x} \le \lambda_{\max}(\mathbf{Q}) \|\mathbf{x}\|_2^2, \qquad \forall \mathbf{x} \in \mathbb{R}^n.$$

Lemma 2.3. Suppose that $u_1, \ldots, u_N$ are real numbers. The arithmetic mean-quadratic mean (AM-QM) inequality, given as part of the mean inequalities in [26, Theorem 2.16], states that

$$\left(\sum_{i=1}^{N} u_i \right)^2 \le N \sum_{i=1}^{N} u_i^2.$$

Proof. The Cauchy–Schwarz inequality for $\mathbf{u} = [u_1, \ldots, u_N]^\mathsf{T} \in \mathbb{R}^N$ and $\mathbf{v} = [1, \ldots, 1]^\mathsf{T} \in \mathbb{R}^N$ states that

$$\left(\sum_{i=1}^{N} u_i \right)^2 = |\langle \mathbf{u}, \mathbf{v} \rangle|^2 \le \|\mathbf{u}\|_2^2 \|\mathbf{v}\|_2^2 = N \sum_{i=1}^{N} u_i^2.$$

□

2.1.3 Definiteness of a Matrix

Let $\mathbf{A} \in \mathbb{S}^n$ be a symmetric matrix. Then $\mathbf{A}$ is said to be:

(i) *Positive definite* if $\mathbf{A} \in \mathbb{S}^n_{++}$, that is, $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} > 0$, $\forall\, \mathbf{x} \neq \mathbf{0}$.

(ii) *Positive semidefinite* if $\mathbf{A} \in \mathbb{S}^n_{+}$, that is, $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \geq 0$, $\forall\, \mathbf{x} \neq \mathbf{0}$.

(iii) *Negative definite* if $\mathbf{A} \in \mathbb{S}^n_{--}$, that is, $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} < 0$, $\forall\, \mathbf{x} \neq \mathbf{0}$.

(iv) *Negative semidefinite* if $\mathbf{A} \in \mathbb{S}^n_{-}$, that is, $\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \leq 0$, $\forall\, \mathbf{x} \neq \mathbf{0}$.

Theorem 2.1 (Principal-Minor Test [27])**.** Given $\mathbf{A} \in \mathbb{S}^n$, a *principal minor of order k* is any determinant obtained by deleting $n-k$ rows and the corresponding $n-k$ columns of $\mathbf{A}$. Then,

1. $\mathbf{A} \in \mathbb{S}^n_{-}$ if and only if every principal minor of even order is nonnegative, and every principal minor of odd order is nonpositive,
2. $\mathbf{A} \in \mathbb{S}^n_{++}$ if and only if every principal minor is positive.

2.2 Signals and Systems

This section reviews the definitions of signals, inner product spaces, systems, and input-output stability.

2.2.1 Signals and Inner Product Spaces

The inputs and outputs of systems are signals, whose properties and relations define the system's input-output characteristics.

Definition 2.4 (Truncation Operator [28])**.** For a discrete-time signal $\mathbf{u} : \mathbb{Z}_{\geq 0} \to \mathbb{R}^n$ and $T \in \mathbb{Z}_{\geq 0}$, its truncation, $\mathbf{u}_T$, is defined as

$$\mathbf{u}_T(t) = \begin{cases} \mathbf{u}(t), & t < T, \\ \mathbf{0}, & t \geq T. \end{cases}$$

Definition 2.5 (Truncated Inner Product [28])**.** For discrete-time signals $\mathbf{u}, \mathbf{y} : \mathbb{Z}_{\geq 0} \to \mathbb{R}^n$, the truncated inner product over the discrete-time interval $\mathcal{T} = \{0, 1, \ldots, T-1\}$ is defined as

$$\langle \mathbf{u}, \mathbf{y} \rangle_T = \langle \mathbf{u}_T, \mathbf{y}_T \rangle = \sum_{t=0}^{\infty} \mathbf{u}_T^\mathsf{T}(t)\mathbf{y}_T(t) = \sum_{t \in \mathcal{T}} \mathbf{u}^\mathsf{T}(t)\mathbf{y}(t), \qquad \forall T \in \mathbb{Z}_{>0}.$$

Definition 2.6 (ℓ_{2e} and ℓ_2 Function Spaces [28])**.** Given a signal $\mathbf{u} : \mathbb{Z}_{\geq 0} \to \mathbb{R}^n$, $\mathbf{u} \in \ell_{2e}$ if

$$\|\mathbf{u}\|_{2T}^2 = \langle \mathbf{u}, \mathbf{u} \rangle_T < \infty, \qquad \forall T \in \mathbb{Z}_{>0}.$$

Moreover, $\mathbf{u} \in \ell_2$ if

$$\|\mathbf{u}\|_2^2 = \langle \mathbf{u}, \mathbf{u} \rangle < \infty.$$

Definition 2.7 (ℓ_∞ Function Space [28])**.** Given a signal $\mathbf{u} : \mathbb{Z}_{\geq 0} \to \mathbb{R}^n$, $\mathbf{u} \in \ell_\infty$ if

$$\|\mathbf{u}\|_\infty = \sup_{t \in \mathbb{Z}_{\geq 0}} \max_{i=1,\ldots,n} |u_i(t)| < \infty.$$

2.2.2 Systems and Input-Output Stability

A *time-invariant* system is a mapping from input signals to output signals such that the output at time t depends only on the input at time t. A linear time-invariant (LTI) System, $\boldsymbol{\mathcal{G}} : \ell_{2e} \to \ell_{2e}$, is a mapping that satisfies

$$\boldsymbol{\mathcal{G}}(a\mathbf{u}_1 + b\mathbf{u}_2) = a\boldsymbol{\mathcal{G}}(\mathbf{u}_1) + b\boldsymbol{\mathcal{G}}(\mathbf{u}_2),$$

for all $\mathbf{u}_1, \mathbf{u}_2 \in \ell_{2e}$ and scalars $a, b \in \mathbb{R}$.

A discrete-time LTI system admits the *state-space* model

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t), \end{aligned} \tag{2.1}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$, $\mathbf{u} \in \mathbb{R}^{n_u}$, and $\mathbf{y} \in \mathbb{R}^{n_y}$ are the state, input, and output signals, respectively, and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ are matrices of appropriate dimensions. The system is *square* if $(n_u = n_y)$. A state-space realization is *minimal* if its state dimension n_x is the smallest possible [29]. The minimal state-space realization of an LTI system $\boldsymbol{\mathcal{G}}$ defined by (2.1) is denoted by

$$\boldsymbol{\mathcal{G}} \overset{\min}{\sim} \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right].$$

The following definition is a special case of the definition of input-output stability for LTI systems.

Definition 2.8 (ℓ_2-Stability [28])**.** The system $\mathbf{y} = \boldsymbol{\mathcal{G}}\mathbf{u}$ with the operator $\boldsymbol{\mathcal{G}} : \ell_{2e} \to \ell_{2e}$ is ℓ_2-stable if $\mathbf{y} \in \ell_2$ for all $\mathbf{u} \in \ell_2$.

Part 1

Passivity of First-Order Optimization Algorithms

Chapter 3

Passivity-Based Control

3.1 Introduction

Passivity is a fundamental concept in control theory that originates from circuit theory, where it can be expressed in terms of resistors, inductors, and capacitors (or, in mechanical systems, masses, springs, and dampers). Intuitively, a system is passive if it does not generate more energy than it absorbs from its inputs, a property that naturally aligns with conservation laws. Notably, passive systems are inherently stable under feedback interconnection, enabling complex systems to be analyzed and designed using energy-based arguments. Nowadays, passivity theory has been extended to the case of nonlinear, time-varying, and distributed systems. This chapter provides the various definitions and theorems related to passivity, which are used throughout this thesis.

3.2 Characterization of Passivity

The notion of passivity, originating from the work of Zames [30] and later formalized within Willems' dissipativity framework [31], is defined as follows.

Definition 3.1 (Passivity [28])**.** Consider a square system with input $\mathbf{u} \in \ell_{2e}$ and output $\mathbf{y} \in \ell_{2e}$ mapped through the operator $\boldsymbol{\mathcal{G}} : \ell_{2e} \to \ell_{2e}$. The system $\boldsymbol{\mathcal{G}}$ is

(i) *passive* if $\exists \beta \in \mathbb{R}_{\leq 0}$ such that

$$\langle \mathbf{u}, \mathbf{y} \rangle_T \geq \beta, \qquad \forall \mathbf{u} \in \ell_{2e}, \forall T \in \mathbb{Z}_{>0},$$

(ii) *input strictly passive (ISP)* if $\exists \beta \in \mathbb{R}_{\leq 0}$ and $\exists \delta \in \mathbb{R}_{>0}$ such that

$$\langle \mathbf{u}, \mathbf{y} \rangle_T \geq \beta + \delta \|\mathbf{u}\|_{2T}^2, \qquad \forall \mathbf{u} \in \ell_{2e}, \forall T \in \mathbb{Z}_{>0},$$

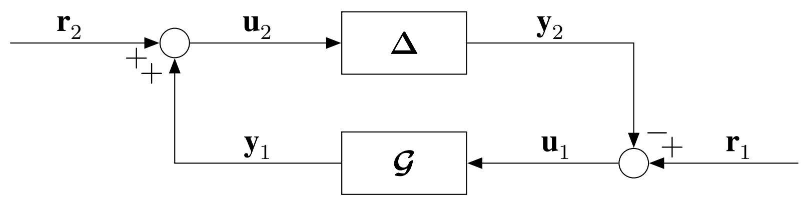


Figure 3.1: The negative feedback interconnection of two systems $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{\Delta}$.

(iii) *output strictly passive (OSP)* if $\exists \beta \in \mathbb{R}_{\leq 0}$ and $\exists \varepsilon \in \mathbb{R}_{>0}$ such that

$$\langle \mathbf{u}, \mathbf{y} \rangle_T \geq \beta + \varepsilon \|\mathbf{y}\|_{2T}^2, \qquad \forall \mathbf{u} \in \ell_{2e}, \forall T \in \mathbb{Z}_{>0},$$

(iv) *very strictly passive (VSP)* if $\exists \beta \in \mathbb{R}_{\leq 0}$ and $\exists \delta, \varepsilon \in \mathbb{R}_{>0}$ such that

$$\langle \mathbf{u}, \mathbf{y} \rangle_T \geq \beta + \delta \|\mathbf{u}\|_{2T}^2 + \varepsilon \|\mathbf{y}\|_{2T}^2, \qquad \forall \mathbf{u} \in \ell_{2e}, \forall T \in \mathbb{Z}_{>0}.$$

The following lemma provides a frequency-domain characterization of passivity.

Lemma 3.1 (Positive Real [32, Lemma 3])**.** Let $\mathbf{G}(z)$ be a square matrix of real rational functions of z and let $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ be a minimal realization of $\mathbf{G}(z)$. Then $\mathbf{G}(z)$ is *positive real* if and only if $\exists \mathbf{P} = \mathbf{P}^\mathsf{T} \succ 0$ such that

$$\begin{bmatrix} \mathbf{A}^\mathsf{T}\mathbf{P}\mathbf{A} - \mathbf{P} & \mathbf{A}^\mathsf{T}\mathbf{P}\mathbf{B} - \mathbf{C}^\mathsf{T} \\ \left(\mathbf{A}^\mathsf{T}\mathbf{P}\mathbf{B} - \mathbf{C}^\mathsf{T}\right)^\mathsf{T} & \mathbf{B}^\mathsf{T}\mathbf{P}\mathbf{B} - \left(\mathbf{D} + \mathbf{D}^\mathsf{T}\right) \end{bmatrix} \preceq 0. \tag{3.1}$$

Remark 3.1. In Lemma 3.1, a necessary and sufficient condition for a positive real transfer matrix is given. However, an LTI system is passive if and only if it has a positive real transfer function [33, Theorem 13.27].

3.3 Passivity Theorems

Having defined notions of passivity in Definition 3.1, the passivity theorem can now be stated.

Theorem 3.1 (Weak Passivity Theorem [28])**.** Consider two systems $\boldsymbol{\mathcal{G}} : \ell_{2e} \to \ell_{2e}$ and $\boldsymbol{\Delta} : \ell_{2e} \to \ell_{2e}$ in negative feedback as per Figure 3.1. Then:

(i) If $\boldsymbol{\mathcal{G}}$ is passive, $\boldsymbol{\Delta}$ is ISP, $\mathbf{r}_1 \in \ell_2$ and $\mathbf{r}_2 = \mathbf{0}$, then $\mathbf{y}_1 \in \ell_2$.

(ii) If $\boldsymbol{\mathcal{G}}$ is passive, $\boldsymbol{\Delta}$ is OSP, $\mathbf{r}_1 = \mathbf{0}$ and $\mathbf{r}_2 \in \ell_2$, then $\mathbf{y}_2 \in \ell_2$.

Notice that the Weak Passivity Theorem can only guarantee the output signals of one of the two systems to be in ℓ_2 at a time. Given that a VSP system is both ISP and OSP, the Strong Passivity Theorem can be used to guarantee the output signals of both systems to be in ℓ_2.

Theorem 3.2 (Strong Passivity Theorem [28])**.** Consider two systems $\boldsymbol{\mathcal{G}} : \ell_{2e} \to \ell_{2e}$ and $\boldsymbol{\Delta} : \ell_{2e} \to \ell_{2e}$ in negative feedback as per Figure 3.1. If $\boldsymbol{\mathcal{G}}$ is passive, $\boldsymbol{\Delta}$ is VSP, and $\mathbf{r}_1, \mathbf{r}_2 \in \ell_2$, then $\mathbf{y}_1, \mathbf{y}_2 \in \ell_2$.

Chapter 4

Input-Output Stability of Gradient Descent: A Passivity-Based Approach

4.1 Overview

This chapter presents a discrete-time passivity-based analysis of the gradient descent method for a class of functions with sector-bounded gradients. Firstly, it is shown that the control interpretation of the gradient descent method, as found in [14], is *not passive*. As such, using a loop transformation, it is shown that the gradient descent method can be interpreted as a passive controller in negative feedback with a very strictly passive system. The passivity theorem is then used to guarantee input-output stability, as well as the global convergence, of the gradient descent method. Furthermore, provided that the lower and upper sector bounds are not equal, the input-output stability of the gradient descent method is guaranteed using the weak passivity theorem for a larger choice of step size.

4.2 Preliminaries

This section provides the necessary background on optimization as a control problem. First, useful function properties such as convexity and smoothness are introduced, along with their generalization to sector-bounded functions. Next, the optimization problem is defined and the gradient descent method is presented. Finally, the control interpretation of the gradient descent method as a Lur'e problem is reviewed.

4.2.1 Function Classes

In order to analyze the stability of optimization algorithms, it is necessary to precisely define the class of functions under consideration. As such, this section outlines the function classes of interest based on properties such as convexity, smoothness, and sector-boundedness.

Definition 4.1 (Convex Function [34])**.** A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ the following inequality holds

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^{\mathsf{T}}(\mathbf{x} - \mathbf{y}). \tag{4.1}$$

Definition 4.2 (m-Strongly Convex Function [34])**.** A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is m-strongly convex if there exists a constant $m \in \mathbb{R}_{>0}$ such that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ the following inequality holds

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^{\mathsf{T}}(\mathbf{x} - \mathbf{y}) + \frac{m}{2}\|\mathbf{x} - \mathbf{y}\|^2. \tag{4.2}$$

Note, f is m-strongly convex if and only if $g(\mathbf{x}) = f(\mathbf{x}) - \frac{m}{2}\|\mathbf{x}\|_2^2$ is convex.

Definition 4.3 (L-Smooth Function [34])**.** A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth if there exists a constant $L \in \mathbb{R}_{>0}$ such that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ the following inequality holds

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))\| \leq L\|\mathbf{x} - \mathbf{y}\|. \tag{4.3}$$

Denote the set of continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ that are m-strongly convex and L-smooth with $0 < m \leq L$ as $\mathcal{F}_{m,L}$. Moreover, $\kappa = L/m$ is called the *condition number* of f. The point $\mathbf{x}^* \in \mathbb{R}^n$ is the *global minimizer* of f if for all $\mathbf{x} \in \mathbb{R}^n$, $f(\mathbf{x}^*) \leq f(\mathbf{x})$. When f is m-strongly convex, $\mathbf{x}^*$ is the unique solution to $\nabla f(\mathbf{x}^*) = 0$.

Definition 4.4 (Co-Coercivity Property [14, Proposition 5])**.** Consider a function $f \in \mathcal{F}_{m,L}$ and its unique global minimizer $\mathbf{x}^*$ such that $\nabla f(\mathbf{x}^*) = \mathbf{0}$. The function $g(\mathbf{x}) = f(\mathbf{x}) - \frac{m}{2}\|\mathbf{x}\|_2^2$ is convex and $(L - m)$–smooth. The co-coercivity of ∇g can be written as

$$\langle \mathbf{x} - \mathbf{y}, \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}) \rangle \geq \frac{mL}{m + L}\|\mathbf{x} - \mathbf{y}\|_2^2 + \frac{1}{m + L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2,$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Notably, for $\mathbf{y} = \mathbf{x}^*$, the inequality reduces to

$$\langle \mathbf{x} - \mathbf{x}^*, \nabla f(\mathbf{x}) \rangle \geq \frac{mL}{m + L}\|\mathbf{x} - \mathbf{x}^*\|_2^2 + \frac{1}{m + L}\|\nabla f(\mathbf{x})\|_2^2. \tag{4.4}$$

Remark 4.1. A useful pair of inequalities that follow from (4.4) are given by

$$m\|\mathbf{x} - \mathbf{x}^*\|_2 \leq \|\nabla f(\mathbf{x})\|_2 \leq L\|\mathbf{x} - \mathbf{x}^*\|_2, \tag{4.5}$$

$$m\|\mathbf{x} - \mathbf{x}^*\|_2^2 \leq \langle \mathbf{x} - \mathbf{x}^*, \nabla f(\mathbf{x}) \rangle \leq L\|\mathbf{x} - \mathbf{x}^*\|_2^2. \tag{4.6}$$

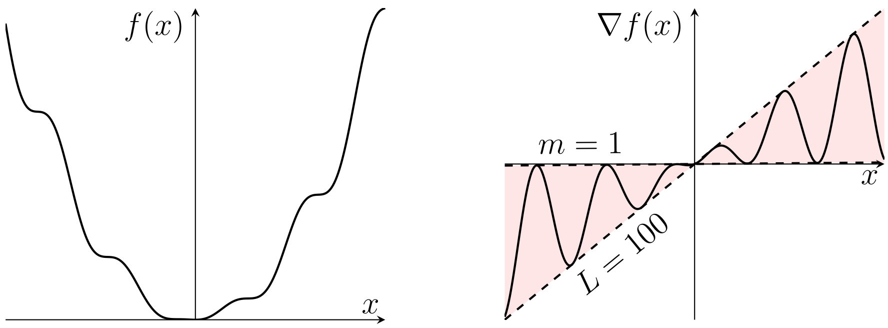


Figure 4.1: The function $f \in \mathcal{S}_{m,L}$ defined in (4.8) for $m = 1$ and $L = 100$. This function is non-convex but has a sector-bounded gradient and a unique global minimizer at $x^* = 0$.

Remark 4.2. If f is convex and L-smooth, the inequality in (4.4) reduces to

$$\langle \mathbf{x} - \mathbf{x}^*, \nabla f(\mathbf{x}) \rangle \geq \frac{1}{L} \|\nabla f(\mathbf{x})\|_2^2. \tag{4.7}$$

The inequality in (4.5) encloses $\nabla f(\mathbf{x})$ in the sector between two lines with slopes m and L [12]. More generally, the set of continuously differentiable functions with unique global minimizers satisfying (4.4) is denoted as $\mathcal{S}_{m,L}$. This class of functions has sector-bounded gradients and contains non-convex functions as well, meaning $\mathcal{F}_{m,L} \subset \mathcal{S}_{m,L}$ [10].

Example 4.1 (A Non-Convex Function with Sector-Bounded Gradient [8]). Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(x) = \frac{L-m}{4}\left(\frac{L+m}{L-m}x^2 + 2\sin(x) - 2x\cos(x)\right), \tag{4.8}$$

for $0 < m < L$. As shown in Figure 4.1, the function f is non-convex, has a unique global minimizer at $x^* = 0$, and $m|x| \leq |\nabla f(x)| \leq L|x|$ for all $x \in \mathbb{R}$.

4.2.2 Optimization Problem

Consider the unconstrained optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}), \tag{4.9}$$

where $f \in \mathcal{S}_{m,L}$. The gradient descent (GD) method with the update rule

$$\mathbf{x}(k+1) = \mathbf{x}(k) - \alpha \nabla f(\mathbf{x}(k)), \tag{4.10}$$

can be used to solve (4.9) using an initialization $\mathbf{x}(0) \in \mathbb{R}^n$ and a constant step size $\alpha \in \mathbb{R}_{>0}$. Note, for the remainder of this chapter, the notation $\mathbf{x}(k)$ is abbreviated to $\mathbf{x}^k$.

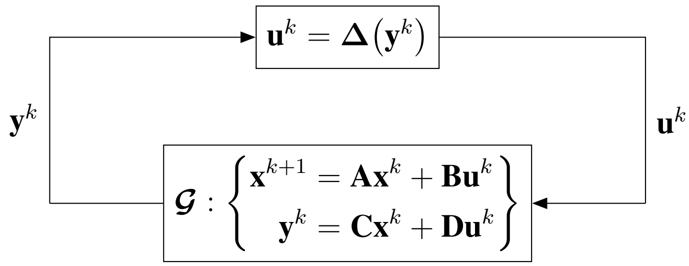


Figure 4.2: The Lur'e problem as an LTI system $\boldsymbol{\mathcal{G}}$ connected in feedback with a static memory-less nonlinearity $\boldsymbol{\Delta}$.

Theorem 4.1 (Convergence of GD [35, Chapter 1.4.2, Theorem 1]). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable and L-smooth function. Then, in the GD algorithm with step size $0 < \alpha < 2/L$, the gradient tends to zero and $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ for all $k \in \mathbb{Z}_{\geq 0}$.

4.2.3 Control Interpretation of Gradient Descent

Consider the Lur'e problem, which is a feedback interconnection of an LTI system $\boldsymbol{\mathcal{G}}$ and a static memoryless nonlinearity $\boldsymbol{\Delta}$ as shown in Figure 3.1. As shown in [14, Section 2], for the special case of $\mathbf{A} = \mathbf{1}$, $\mathbf{B} = -\alpha\mathbf{1}$, $\mathbf{C} = \mathbf{1}$, $\mathbf{D} = \mathbf{0}$, and $\boldsymbol{\Delta} = \nabla f$, it follows that

$$\begin{cases} \mathbf{x}^{k+1} = \mathbf{x}^k - \alpha \mathbf{u}^k, \\ \mathbf{y}^k = \mathbf{x}^k, \\ \mathbf{u}^k = \nabla f(\mathbf{y}^k). \end{cases} \tag{4.11}$$

The closed-loop dynamics of (4.11) is described by $\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha\nabla f(\mathbf{x}^k)$, which is equivalent to the GD update rule in (4.10). As such, existing control theory results regarding the stability of Lur'e problems can be applied to analyze the stability of the GD method. Importantly, within the context of this framework, stability refers to the ℓ_2-stability of the interconnected system, as defined in Definition 2.8.

4.3 Input-Output Stability of Gradient Descent

By modifying the approach in Section 4.2.3, the GD method can be reformulated as the *negative* feedback interconnection shown in Figure 3.1, where $\boldsymbol{\Delta} = \nabla f$ and $\boldsymbol{\mathcal{G}}$ is an LTI system with minimal realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = (\mathbf{1}, \alpha\mathbf{1}, \mathbf{1}, \mathbf{0})$. Moreover, for the general case of $\mathbf{x}^* \neq 0$, $\boldsymbol{\Delta} = \nabla f$ is not a bounded operator as it does not map zero inputs to zero outputs.

Remark 4.3. Similar to [10], a state shifting argument is used such that $\boldsymbol{\Delta}$ in Figure 3.1 maps $\mathbf{u}_2$ to $\mathbf{y}_2$ as $\mathbf{y}_2^k = \nabla f(\mathbf{u}_2^k + \mathbf{x}^*)$. It follows that for $\mathbf{u}_2^k = \mathbf{0}$, $\mathbf{y}_2^k = \nabla f(\mathbf{x}^*) = \mathbf{0}$. Moreover,

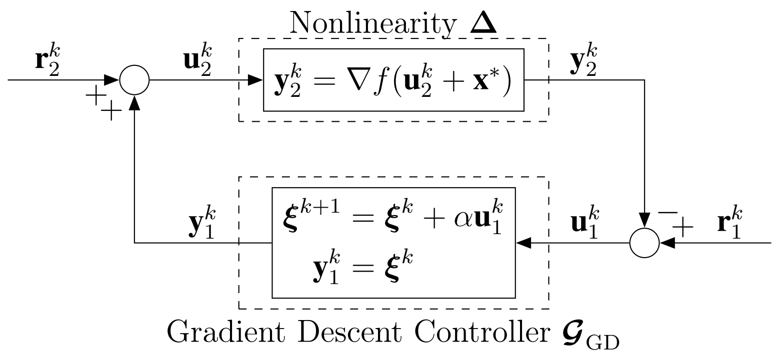


Figure 4.3: The negative feedback interconnection of the GD controller, $\boldsymbol{\mathcal{G}}_{\text{GD}}$, and the nonlinearity $\boldsymbol{\Delta}$, being the shifted gradient that maps zero inputs to zero outputs.

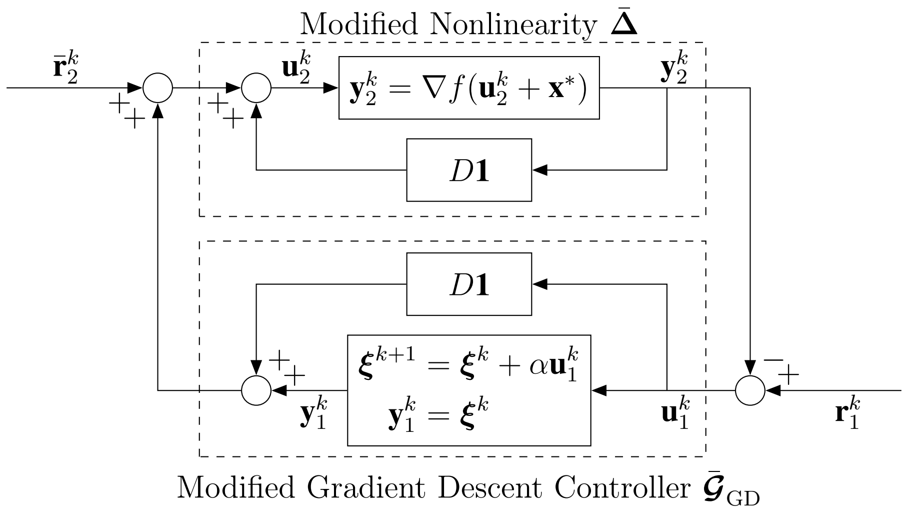


Figure 4.4: Loop transformation of the negative feedback interconnection representation of the GD method in Figure 4.3. The loop transformation introduces a feedthrough term $D\mathbf{1}$, resulting in the modified GD controller $\bar{\boldsymbol{\mathcal{G}}}_{\text{GD}}$, the modified nonlinearity $\bar{\boldsymbol{\Delta}}$, and $\bar{\mathbf{r}}_2^k = \mathbf{r}_2^k - D\mathbf{r}_1^k$.

using the shifted state $\boldsymbol{\xi}^k = \mathbf{x}^k - \mathbf{x}^*$, the system $\boldsymbol{\mathcal{G}}$ in Figure 3.1 can be represented as

$$\boldsymbol{\xi}^{k+1} = \mathbf{A}\boldsymbol{\xi}^k + \mathbf{B}\mathbf{u}_1^k, \tag{4.12a}$$

$$\mathbf{y}_1^k = \mathbf{C}\boldsymbol{\xi}^k + \mathbf{D}\mathbf{u}_1^k. \tag{4.12b}$$

Finally, the GD update rule in (4.10) can then be written in the form of (4.12) with minimal realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = (\mathbf{1}, \alpha\mathbf{1}, \mathbf{1}, \mathbf{0})$ as shown in Figure 4.3. Additionally, the convergence rate of the GD method becomes analogous to the rate at which $\mathbf{y}_1^k = \boldsymbol{\xi}^k \to \mathbf{0}$.

4.3.1 Loop Transformation

The GD controller $\boldsymbol{\mathcal{G}}_{\mathrm{GD}}$, shown in Figure 4.3, is *strictly proper*, since $\mathbf{D} = \mathbf{0}$. However, a discrete-time system of the form (4.12) with $\mathbf{D} = \mathbf{0}$ can never be passive [36, Remark 2.7]. Therefore, to analyze the passivity of $\boldsymbol{\mathcal{G}}_{\mathrm{GD}}$, the loop transformation in Figure 4.4 is performed, effectively introducing a constant feedthrough term $\mathbf{D} = D\mathbf{1}$, where $D \in \mathbb{R}_{>0}$. This feedthrough term results in the modified GD controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ with a minimal realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = (\mathbf{1}, \alpha\mathbf{1}, \mathbf{1}, D\mathbf{1})$.

Remark 4.4. This loop transformation does not change the closed-loop dynamics of the system, meaning Figure 4.4 still represents the GD method with update rule (4.10).

Within the context of convex constrained optimization problems, similar loop transformation techniques have been used in [11, 37] to analyze the stability of discrete-time alternating direction method of multipliers (ADMM) and primal-dual algorithms, respectively.

To guarantee the input-output stability of the negative feedback interconnection in Figure 4.4 using the passivity theorem, it is sufficient to find conditions under which the modified GD controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ is passive and the modified nonlinearity $\bar{\boldsymbol{\Delta}}$ is VSP.

4.3.2 Passivity of the Modified Gradient Descent Controller

In this subsection, the sufficient amount of the feedthrough term D required to render $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ passive is shown to be directly related to the step size α.

Lemma 4.1 ([1, Lemma 2])**.** The modified GD controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ in Figure 4.4 is passive if and only if $D \geq \alpha/2 > 0$.

Proof. Substituting the minimal realization of $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ into (3.1) yields

$$\mathbf{M} = \begin{bmatrix} \mathbf{0} & \alpha\mathbf{P} - \mathbf{1} \\ \alpha\mathbf{P} - \mathbf{1} & \alpha^2\mathbf{P} - 2D\mathbf{1} \end{bmatrix}.$$

From Theorem 2.1, $\mathbf{M} \preceq 0$ if and only if

$$\alpha\mathbf{P} - \mathbf{1} = \mathbf{0}, \qquad \alpha^2\mathbf{P} - 2D\mathbf{1} \preceq 0. \tag{4.13}$$

From (4.13), it follows that $\mathbf{P} = \frac{1}{\alpha}\mathbf{1}$ and $D \geq \alpha/2$, for $\alpha \in \mathbb{R}_{>0}$. Finally, Lemma 3.1 in tandem with Remark 3.1 concludes that $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ is passive if and only if $D \geq \alpha/2$, for $\alpha \in \mathbb{R}_{>0}$. □

4.3.3 Passivity of the Modified Nonlinearity

In this subsection, an upper bound on the feedback term D is derived such that the modified nonlinearity $\bar{\boldsymbol{\Delta}}$ has certain passivity properties. Firstly, it is shown that for $f \in \mathcal{S}_{m,L}$, the

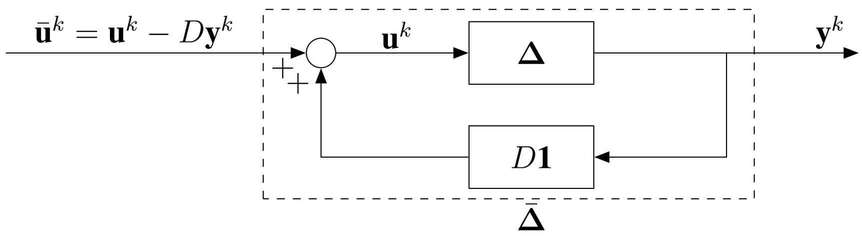


Figure 4.5: A discrete-time system $\bar{\boldsymbol{\Delta}}$, composed of the positive feedback interconnection of a VSP system $\boldsymbol{\Delta}$, and the feedback term $D\mathbf{1}$.

sector-bound condition in (4.4) leads to $\boldsymbol{\Delta}$ being VSP.

Lemma 4.2 ([1, Lemma 3]). Consider a function $f \in \mathcal{S}_{m,L}$ with $m, L \in \mathbb{R}_{>0}$, such that $m \leq L$, and its unique global minimizer $\mathbf{x}^*$. The operator $\boldsymbol{\Delta}$ in Figure 4.3 mapping $\mathbf{u}_2^k$ to $\mathbf{y}_2^k$ as $\mathbf{y}_2^k = \nabla f(\mathbf{u}_2^k + \mathbf{x}^*)$ is VSP with

$$\varepsilon = \frac{1}{m+L} > 0, \qquad \delta = \frac{mL}{m+L} > 0. \tag{4.14}$$

Proof. For a given discrete time interval $\mathcal{T} = \{0, 1, \ldots, T-1\}$, with $T \in \mathbb{Z}_{>0}$, consider the iterates $\{\mathbf{y}_2^k\}$ obtained from the operator $\boldsymbol{\Delta}$, such that $\mathbf{y}_2^k = \nabla f(\mathbf{u}_2^k + \mathbf{x}^*)$. For $f \in \mathcal{S}_{m,L}$, the sector-bound inequality in (4.4) can be written as

$$\langle \mathbf{u}_2^k, \mathbf{y}_2^k \rangle \geq \frac{mL}{m+L} \|\mathbf{u}_2^k\|_2^2 + \frac{1}{m+L} \|\mathbf{y}_2^k\|_2^2, \tag{4.15}$$

for all $\mathbf{u}_2^k \in \mathbb{R}^n$ and $k \in \mathcal{T}$. Summing (4.15) over $k \in \mathcal{T}$ and defining ε and δ as per (4.14), it follows that

$$\sum_{k \in \mathcal{T}} \langle \mathbf{u}_2^k, \mathbf{y}_2^k \rangle = \langle \mathbf{y}_2, \mathbf{u}_2 \rangle_T \geq \delta \|\mathbf{u}_2\|_{2T}^2 + \varepsilon \|\mathbf{y}_2\|_{2T}^2,$$

for all $\mathbf{u} \in \ell_{2e}$ and $T \in \mathbb{Z}_{>0}$. Therefore, $\boldsymbol{\Delta}$ is VSP with $\varepsilon, \delta \in \mathbb{R}_{>0}$ and $\beta = 0$. $\square$

Remark 4.5. Using a similar approach as in Lemma 4.2, it follows from Equation (4.7) that for a function $f \in \mathcal{S}_{0,L}$, the operator $\boldsymbol{\Delta}$ is OSP with

$$\varepsilon = \frac{1}{L} > 0. \tag{4.16}$$

What remains is to derive an upper bound on D to ensure the positive feedback interconnection of $\boldsymbol{\Delta}$ and $D\mathbf{1}$ as per Figure 4.5 results in a system $\bar{\boldsymbol{\Delta}}$ with certain passivity properties.

Lemma 4.3. [1, Lemma 4] Consider a VSP system $\boldsymbol{\Delta}$ with ε and δ in (4.14) and $\beta = 0$.

Given a feedback term $D\mathbf{1}$ with $D \in \mathbb{R}_{>0}$, if $D < 1/L$, the positive feedback interconnection of $\boldsymbol{\Delta}$ and $D\mathbf{1}$ results in a VSP system $\bar{\boldsymbol{\Delta}}$.

Proof. From the VSP property of $\boldsymbol{\Delta}$, it follows that

$$\langle \mathbf{u}, \mathbf{y} \rangle_T \geq \delta \|\mathbf{u}\|_{2T}^2 + \varepsilon \|\mathbf{y}\|_{2T}^2, \qquad \forall \mathbf{u} \in \ell_{2e}, \forall T \in \mathbb{Z}_{>0}. \tag{4.17}$$

Adding and subtracting the terms $\delta D^2 \|\mathbf{y}\|_{2T}^2$ and $2\delta D \langle \mathbf{u}, \mathbf{y} \rangle_T$ from (4.17) and completing the square, it follows that

$$\langle \mathbf{u}, \mathbf{y} \rangle_T \geq \delta \|\bar{\mathbf{u}}\|_{2T}^2 + \left(\varepsilon - \delta D^2\right) \|\mathbf{y}\|_{2T}^2 + 2\delta D \langle \mathbf{u}, \mathbf{y} \rangle_T, \tag{4.18}$$

where $\bar{\mathbf{u}}^k = \mathbf{u}^k - D\mathbf{y}^k$. Rearranging (4.18) and collecting similar terms yields

$$(1 - 2\delta D) \langle \mathbf{u}, \mathbf{y} \rangle_T \geq \delta \|\bar{\mathbf{u}}\|_{2T}^2 + \left(\varepsilon - \delta D^2\right) \|\mathbf{y}\|_{2T}^2.$$

For $1 - 2\delta D > 0$, it follows that

$$\langle \mathbf{u}, \mathbf{y} \rangle_T \geq \frac{\delta}{1 - 2\delta D} \|\bar{\mathbf{u}}\|_{2T}^2 + \frac{(\varepsilon - \delta D^2)}{1 - 2\delta D} \|\mathbf{y}\|_{2T}^2. \tag{4.19}$$

Subtracting $D\|\mathbf{y}\|_{2T}^2$ from both sides of (4.19) and combining similar terms results in

$$\begin{aligned} \langle \bar{\mathbf{u}}, \mathbf{y} \rangle_T &\geq \frac{\delta}{1 - 2\delta D} \|\bar{\mathbf{u}}\|_{2T}^2 + \frac{(\varepsilon - D + \delta D^2)}{1 - 2\delta D} \|\mathbf{y}\|_{2T}^2 \\ &= \bar{\delta} \|\bar{\mathbf{u}}\|_{2T}^2 + \bar{\varepsilon} \|\mathbf{y}\|_{2T}^2, \end{aligned}$$

where $\bar{\delta} = \delta/(1 - 2\delta D)$ and $\bar{\varepsilon} = (\varepsilon - D + \delta D^2)/(1 - 2\delta D)$. For $\bar{\boldsymbol{\Delta}}$ to be VSP with $\bar{\varepsilon}, \bar{\delta} \in \mathbb{R}_{>0}$, it is required that

$$D < \frac{1}{2\delta}, \qquad\qquad 0 < \varepsilon - D + \delta D^2. \tag{4.20}$$

Note, $D^* = 1/(2\delta)$ is the unique global minimum of the quadratic equation $0 = \varepsilon - D + \delta D^2$. Substituting $\varepsilon = 1/(m + L)$ and $\delta = mL/(m + L)$ into (4.20), it follows that

$$D < \frac{m + L}{2mL}, \qquad\qquad 0 < \frac{1}{m + L} - D + \frac{mL}{m + L} D^2. \tag{4.21}$$

The roots of the quadratic equation (4.21) are $r_1 = 1/L$ and $r_2 = 1/m$. For $m \leq L$, it follows that $r_1 \leq D^* \leq r_2$, where equality is achieved for $m = L$. For $D < r_1 = 1/L$, both inequalities in (4.21) are satisfied, and therefore, the positive feedback interconnection of $\boldsymbol{\Delta}$

and $D\mathbf{1}$ results in a VSP system $\bar{\mathbf{\Delta}}$, with $\bar{\varepsilon}, \bar{\delta} \in \mathbb{R}_{>0}$. $\square$

Lemma 4.4. [1, Lemma 5] Consider a VSP system $\mathbf{\Delta}$ with ε and δ in (4.14) and $\beta = 0$. Given a feedback term $D\mathbf{1}$ with $D \in \mathbb{R}_{>0}$, if $D = 1/L$ and $m < L$, the positive feedback interconnection of $\mathbf{\Delta}$ and $D\mathbf{1}$ results in an ISP system $\bar{\mathbf{\Delta}}$.

Proof. The proof follows similar to Lemma 4.3, with the only difference being that for $\bar{\mathbf{\Delta}}$ to be ISP, with $\bar{\delta} \in \mathbb{R}_{>0}$ and $\bar{\varepsilon} = 0$, it is required that

$$D < \frac{m+L}{2mL}, \qquad 0 = \frac{1}{m+L} - D + \frac{mL}{m+L} D^2. \tag{4.22}$$

As such, for $D = r_1 = 1/L$, both conditions in (4.22) are satisfied. $\square$

Note, Lemma 4.3 and Lemma 4.4 consider the case of $\mathbf{\Delta}$ being VSP. Herein, a less restrictive condition is derived for the case of $\mathbf{\Delta}$ being OSP.

Lemma 4.5. Consider an OSP system $\mathbf{\Delta}$ with $\varepsilon = 1/L$ and $\beta = 0$. Given a feedback term $D\mathbf{1}$ with $D \in \mathbb{R}_{>0}$, if $D < 1/L$, the positive feedback interconnection of $\mathbf{\Delta}$ and $D\mathbf{1}$ results in an OSP system $\bar{\mathbf{\Delta}}$.

Proof. From the OSP property of $\mathbf{\Delta}$, it follows that

$$\langle \mathbf{u}, \mathbf{y} \rangle_T \geq \varepsilon \|\mathbf{y}\|_{2T}^2, \qquad \forall \mathbf{u} \in \ell_{2e}, \forall T \in \mathbb{Z}_{>0}. \tag{4.23}$$

Subtracting the term $D\|\mathbf{y}\|_{2T}^2$ from both sides of (4.23), it follows that

$$\langle \mathbf{u}, \mathbf{y} \rangle_T - D\|\mathbf{y}\|_{2T}^2 = \langle \bar{\mathbf{u}}, \mathbf{y} \rangle_T \geq (\varepsilon - D)\|\mathbf{y}\|_{2T}^2. \tag{4.24}$$

As such, for $D < \varepsilon$, the system $\bar{\mathbf{\Delta}}$ is OSP. $\square$

4.3.4 Input-Output Stability the Negative Feedback Interconnection

In Section 4.3.2 the sufficient amount of the feedthrough term D required to render $\bar{\boldsymbol{\mathcal{G}}}_{\text{GD}}$ passive was shown to be directly related to the step size α. In Section 4.3.3, an upper bound on D was derived such that the modified nonlinearity $\bar{\mathbf{\Delta}}$ has certain passivity properties. Finally, in this subsection, the input-output stability of the negative feedback interconnection in Figure 4.4 is established using the passivity theorem.

Theorem 4.2 ([1, Theorem 3])**.** Consider the negative feedback interconnection representation of the GD method shown in Figure 4.4, where $f \in \mathcal{S}_{m,L}$ with $0 < m \leq L$, and $D = \alpha/2$.

(i) Provided the step size $\alpha \in (0, \frac{2}{L})$ and $\mathbf{r}_1, \bar{\mathbf{r}}_2 \in \ell_2$, then $\mathbf{y}_1 + D\mathbf{u}_1 \in \ell_2$ and $\mathbf{y}_2 \in \ell_2$.

(*ii*) Additionally, for $m < L$, provided the step size $\alpha \in \left(0, \frac{2}{L}\right]$, $\bar{\mathbf{r}}_2 = \mathbf{0}$, and $\mathbf{r}_1 \in \ell_2$, then $\mathbf{y}_1 + D\mathbf{u}_1 \in \ell_2$.

Proof. As per Lemma 4.1, $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ is passive if and only if $0 < \alpha/2 \leq D$. Moreover, for $f \in \mathcal{S}_{m,L}$, provided $D < 1/L$, $\bar{\boldsymbol{\Delta}}$ is VSP, as per Lemma 4.3. It follows that for $D = \alpha/2$ with $\alpha \in (0, 2/L)$, $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ is passive and $\bar{\boldsymbol{\Delta}}$ is VSP. The proof of (*i*) then follows from Theorem 3.2. Furthermore, assuming $m < L$ and $D = \alpha/2 = 1/L$, $\bar{\boldsymbol{\Delta}}$ is ISP, as per Lemma 4.4. Consequently, the proof of (*ii*) then follows from Theorem 3.1-(i), where for $\alpha = 2/L$, $\bar{\boldsymbol{\Delta}}$ is ISP and $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ is passive. □

Note, Theorem 4.2 establishes the input-output stability of the gradient descent algorithm for the case of $f \in \mathcal{S}_{m,L}$ with $0 < m \leq L$. Herein, a similar result is derived for the case of $f \in \mathcal{S}_{0,L}$ with $L > 0$.

Theorem 4.3. Consider the negative feedback interconnection representation of the GD method shown in Figure 4.4, where $f \in \mathcal{S}_{0,L}$ with $0 < L$, and $D = \alpha/2$. Provided the step size $\alpha \in (0, \frac{2}{L})$, $\mathbf{r}_1 = \mathbf{0}$, and $\bar{\mathbf{r}}_2 \in \ell_2$, then $\mathbf{y}_2 \in \ell_2$.

Proof. As per Lemma 4.5, provided $D < 1/L$, $\bar{\boldsymbol{\Delta}}$ is OSP. The result then follows from Theorem 3.1-(ii), where for $\alpha \in (0, 2/L)$, $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$ is passive and $\bar{\boldsymbol{\Delta}}$ is OSP. □

4.4 Discussion

The passivity-based analysis presented in Section 4.3 decouples the stability analysis of the GD controller from the properties of the objective function. Notice, Lemma 4.1 provides a lower bound on the feedthrough term D and dictates the necessary and sufficient condition for the controller to be passive. On the other hand, using the sector-bounds of the nonlinearity $\boldsymbol{\Delta}$, Lemmas 4.2, 4.4 and 4.5 provide an upper bound on the feedback term D. Therefore, for a function $f \in \mathcal{S}_{m,L}$ with $0 < m \leq L$, the passivity-based analysis of other first-order optimization algorithms can be approached as follows:

1. Cast the algorithm as an LTI controller in negative feedback with a static memoryless nonlinearity as per Figure 4.3.
2. Using the minimal realization of the controller, determine the lower bound on the feedthrough term D by solving the linear matrix inequality (LMI) in (3.1).
3. Determine the bounds on the algorithm's step size(s) that guarantee the feedthrough term D is within the bounds provided by Lemmas 4.3 and 4.4. Consequently, use the appropriate passivity theorem to guarantee the input-output stability of the algorithm.

Table 4.1: Summary of the Input-Output Stability Theorems for the GD Method.

	Function Class	Sector Bounds	Condition	Guarantee
Theorem 4.2-(i)	$f \in \mathcal{S}_{m,L}$	$0 < m \leq L$	$\alpha \in (0, \frac{2}{L})$	(4.25) & (4.26)
Theorem 4.2-(ii)	$f \in \mathcal{S}_{m,L}$	$0 < m < L$	$\alpha = \frac{2}{L}$	(4.26)
Theorem 4.3	$f \in \mathcal{S}_{0,L}$	$0 < L$	$\alpha \in (0, \frac{2}{L})$	(4.25)

4.4.1 Interpretation of Input-Output Stability

To draw a connection between the input-output stability of the negative feedback interconnection in Figure 4.4 and the convergence of the GD method in (4.10), it is assumed that $\mathbf{r}_1^k = \mathbf{r}_2^k = \mathbf{0}$ for all $k \in \mathbb{Z}_{\geq 0}$. Moreover, the following useful property of ℓ_2 signals is used.

Remark 4.6. If $\boldsymbol{\xi} \in \ell_2$, then $\boldsymbol{\xi}^k \to \mathbf{0}$ as $k \to \infty$ [28].

Consider the negative feedback interconnection in Figure 4.4 where $\mathbf{y}_2^k$ is the output of the modified nonlinearity $\bar{\boldsymbol{\Delta}}$ and $\mathbf{y}_1^k + D\mathbf{u}_1^k$ is the output of the modified GD controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$. If $\mathbf{y}_2 \in \ell_2$ it follows that

$$\mathbf{y}_2^k = \nabla f(\boldsymbol{\xi}^k + \mathbf{x}^*) = \nabla f(\mathbf{x}^k) \to \mathbf{0}. \tag{4.25}$$

Similarly, $\mathbf{y}_1 + D\mathbf{u}_1 \in \ell_2$ implies that

$$\mathbf{x}^k - D\nabla f(\mathbf{x}^k) \to \mathbf{x}^*. \tag{4.26}$$

Note (4.25) implies that $\mathbf{x}^k \to \mathbf{x}^*$, since $\mathbf{x}^*$ is assumed to be the unique point such that $\nabla f(\mathbf{x}^*) = \mathbf{0}$. By contrast, (4.26) alone guarantees neither $\mathbf{x}^k \to \mathbf{x}^*$ nor $\nabla f(\mathbf{x}^k) \to \mathbf{0}$ as $k \to \infty$. Using the relations (4.25) and (4.26), the passivity-based analysis of the GD method presented in Section 4.3 is summarized in Table 4.1.

Firstly, both Theorem 4.2-(i) and Theorem 4.3 guarantee the global convergence of the GD method for $\alpha \in (0, \frac{2}{L})$ via (4.25). However, unlike Theorem 4.2-(i), Theorem 4.3 does not require knowledge of the lower sector bound m. Therefore, Theorem 4.3 can guarantee the global convergence of the GD method for f being convex and L-smooth, whereas Theorem 4.2-(i) further requires f to be m-strongly convex. This result is consistent with Polyak's convergence analysis of the GD method in Theorem 4.1, where f is only assumed to be L-smooth and bounded from below. For $f \in \mathcal{F}_{m,L}$, this convergence region is also obtained within the framework of equilibrium-independent dissipativity (EID) with quadratic supply rates in [38, Example 5.4].

Secondly, Theorem 4.2-(ii) considers the case where $\alpha = \frac{2}{L}$ and can only guarantee (4.26). The example below highlights the significance of this result.

Example 4.2. Inspired by the counter example provided in [35], consider the objective function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}$, where $\mathbf{x} = [x_1 \quad x_2]^\mathsf{T}$, $\mathbf{A} = \text{diag}(m, L)$, and $m < L$. It can be shown that $f \in \mathcal{F}_{m,L}$, and therefore, $f \in \mathcal{S}_{m,L}$. Given $\mathbf{x}^0 \in \mathbb{R}^2$, the first iteration of the GD method in (4.10) with $\alpha = 2/L$ yields

$$\mathbf{x}^1 = \mathbf{x}^0 - \frac{2}{L}\nabla f(\mathbf{x}^0) = \mathbf{x}^0 - \frac{2}{L}\begin{bmatrix} m & 0 \\ 0 & L \end{bmatrix}\mathbf{x}^0 = \begin{bmatrix} \frac{L-2m}{L}x_1^0 \\ -x_2^0 \end{bmatrix}.$$

Notice $x_2^{k+1} = (-1)^{k+1}x_2^k$, for $k \in \mathbb{Z}_{\geq 0}$, signifying the oscillatory behavior of the series of iterations $\{x_2^k\}$. As such, the GD method with $\alpha = 2/L$ does converge to the optimal solution $x_2^* = 0$. However, given that $D = \alpha/2 = 1/L$, (4.26) still holds since $x_2^k - (1/L)(Lx_2^k) = x_2^* = 0$. This means that (4.26) can still be used to obtain the optimal solution x_2^* even though the series of iterations $\{x_2^k\}$ does not converge to x_2^*.

As shown in Example 4.2, (4.26) can be used to obtain the optimal solution even when the series of iterations does not converge. As such, (4.26) can be interpreted as an additional stopping criterion for the GD method. To this end, it follows from (4.26) that

$$\left(\mathbf{x}^k - D\nabla f(\mathbf{x}^k)\right) - \left(\mathbf{x}^{k-1} - D\nabla f(\mathbf{x}^{k-1})\right) \to \mathbf{0}. \tag{4.27}$$

Since $D = \alpha/2$, substituting the GD update rule (4.10) into (4.27) and simplifying yields $\nabla f(\mathbf{x}^k) + \nabla f(\mathbf{x}^{k-1}) \to \mathbf{0}$. As such, in practice, the stopping criterion

$$\left\|\nabla f(\mathbf{x}^k) + \nabla f(\mathbf{x}^{k-1})\right\|_2^2 < \epsilon,$$

for an arbitrarily small $\epsilon \in \mathbb{R}_{>0}$, can be used to verify the relation in (4.26).

4.5 Closing Remarks

This chapter presented a discrete-time passivity-based analysis of the GD method. For an objective function with sector-bounded gradient and a unique global minimizer, $f \in \mathcal{S}_{m,L}$, it is shown that the shifted gradient $\boldsymbol{\Delta}$ is VSP for $0 < m$ and OSP for $m = 0$. Using a loop transformation, it is shown that the GD method can be interpreted as a passive controller in negative feedback with a VSP system, provided $\alpha \in (0, 2/L)$. The input-output stability, as well as the global convergence, of the GD method is then guaranteed using the strong version of the passivity theorem. Moreover, this chapter relaxes the assumptions in [1] by guaranteeing the global convergence of the GD method for $f \in \mathcal{S}_{0,L}$ with $L > 0$ and $\alpha \in (0, 2/L)$. Furthermore, provided $m < L$, it is shown that for $\alpha = 2/L$, the passive GD controller is now in negative feedback with an ISP system instead. Therefore, the weak passivity theorem is used to guarantee the input-output stability of the GD method. For this larger choice of step size, Polyak's counter example in [35] is revisited, and a new stopping criterion is proposed to ensure the GD method can still be used to obtain the unique global minimizer.

Chapter 5

Input-Output Stability of Momentum-Based Methods: A Passivity-Based Approach

5.1 Overview

This chapter extends the passivity-based analysis of the gradient descent method presented in Chapter 4 to popular first-order momentum-based methods. For each algorithm, a loop transformation is performed, resulting in the corresponding modified momentum-based controllers, $\bar{\boldsymbol{\mathcal{G}}}$. The main objective of this chapter is to determine the lower bound on the feedthrough term of $\bar{\boldsymbol{\mathcal{G}}}$ to ensure its passivity. Importantly, for $f \in \mathcal{S}_{0,L}$, the analysis for the upper bound on the feedthrough term remains unchanged. As such, the proposed passivity-based analysis guarantees the global convergence of the momentum-based methods for a broader class of functions than those considered in the literature.

5.2 Preliminaries

This section reviews popular first-order momentum-based methods and their control interpretation. Over the years, many variations of the GD method have been proposed to improve its convergence rate when solving the unconstrained optimization problem in (4.9). One such variation are the momentum-based methods, which incorporate a *momentum* terms to accelerate convergence. Examples of famous momentum-based methods include Polyak's heavy ball (HB) method [6],

$$\mathbf{x}^{k+1} = (1+\beta)\mathbf{x}^k - \beta\mathbf{x}^{k-1} - \alpha\nabla f(\mathbf{x}^k), \tag{5.1}$$

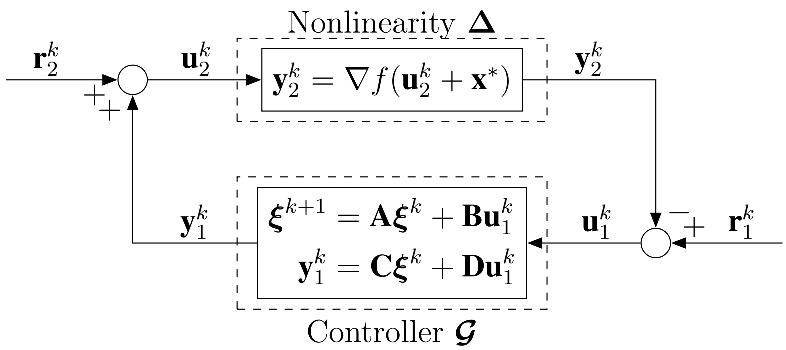


Figure 5.1: The negative feedback interconnection of the controller, $\boldsymbol{\mathcal{G}}$, and the nonlinearity $\boldsymbol{\Delta}$, being the shifted gradient that maps zero inputs to zero outputs.

and Nesterov's accelerated gradient (NAG) method [7],

$$\begin{aligned} \mathbf{x}^{k+1} &= \mathbf{y}^k - \alpha \nabla f(\mathbf{y}^k), \\ \mathbf{y}^k &= (1+\beta)\mathbf{x}^k - \beta \mathbf{x}^{k-1}, \end{aligned} \tag{5.2}$$

where $\alpha, \beta \in \mathbb{R}_{>0}$ are the step size and momentum parameters, respectively.

More generally, as discussed in [15], many momentum-based methods can be written using the recursion

$$\begin{aligned} \mathbf{x}^{k+1} &= (1+\beta)\mathbf{x}^k - \beta \mathbf{x}^{k-1} - \alpha \nabla f(\mathbf{y}^k), \\ \mathbf{y}^k &= (1+\gamma)\mathbf{x}^k - \gamma \mathbf{x}^{k-1}, \\ \boldsymbol{\zeta}^k &= (1+\delta)\mathbf{x}^k - \delta \mathbf{x}^{k-1}, \end{aligned} \tag{5.3}$$

where $\mathbf{x} \in \ell_{2e}$ is the internal state, the gradient is applied to the intermediate variable $\mathbf{y} \in \ell_{2e}$, and the output is $\boldsymbol{\zeta} \in \ell_{2e}$. The parameters $\alpha \in \mathbb{R}_{>0}$ is the step size, while $\beta, \gamma, \delta \in \mathbb{R}_{\geq 0}$ are the momentum parameters.

For $f \in \mathcal{F}_{m,L}$, where $\kappa = L/m$ is the condition number, the triple momentum (TM) method [15] is defined by (5.3) with

$$(\alpha, \beta, \gamma, \delta) = \left(\frac{1+\rho}{L}, \frac{\rho^2}{2-\rho}, \frac{\rho^2}{(1+\rho)(2-\rho)}, \frac{\rho^2}{1-\rho^2} \right), \tag{5.4}$$

where $\rho = 1 - 1/\sqrt{\kappa}$. Note, $\boldsymbol{\zeta}$ is only an output mapping of the internal state $\mathbf{x}$ and does not affect the recursion. Therefore, the remainder of this chapter focuses on the reduced two-state system in $\mathbf{x}$ and $\mathbf{y}$.

Table 5.1: Minimal Realization of $\boldsymbol{\mathcal{G}}$ for Momentum-Based Methods.

$\boldsymbol{\mathcal{G}}_{\text{HB}}$	$\boldsymbol{\mathcal{G}}_{\text{NAG}}$	$\boldsymbol{\mathcal{G}}_{\text{TM}}$
$\left[\begin{array}{cc\|c} (1+\beta)\mathbf{1} & -\beta\mathbf{1} & \alpha\mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{1} & \mathbf{0} & \mathbf{0} \end{array}\right]$	$\left[\begin{array}{cc\|c} (1+\beta)\mathbf{1} & -\beta\mathbf{1} & \alpha\mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \hline (1+\beta)\mathbf{1} & -\beta\mathbf{1} & \mathbf{0} \end{array}\right]$	$\left[\begin{array}{cc\|c} (1+\beta)\mathbf{1} & -\beta\mathbf{1} & \alpha\mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \hline (1+\gamma)\mathbf{1} & -\gamma\mathbf{1} & \mathbf{0} \end{array}\right]$

5.2.1 Control Interpretation of Momentum-Based Methods

To express the momentum-based methods introduced in this section using the negative feedback interconnection in Figure 5.1, consider the augmented state

$$\boldsymbol{\xi}^k = \begin{bmatrix} \mathbf{x}^k - \mathbf{x}^* \\ \mathbf{x}^{k-1} - \mathbf{x}^* \end{bmatrix}, \tag{5.5}$$

where $\mathbf{x}^*$ is the unique minimizer of f. As such, the recursion in (5.3) can be written as

$$\begin{aligned} \boldsymbol{\xi}^{k+1} &= \begin{bmatrix} (1+\beta)\mathbf{1} & -\beta\mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \boldsymbol{\xi}^k + \begin{bmatrix} \alpha\mathbf{1} \\ \mathbf{0} \end{bmatrix} \mathbf{u}_1^k, \\ \mathbf{y}_1^k &= \begin{bmatrix} (1+\gamma)\mathbf{1} & -\gamma\mathbf{1} \end{bmatrix} \boldsymbol{\xi}^k, \end{aligned} \tag{5.6}$$

Following (5.6), the minimal realization of the controller $\boldsymbol{\mathcal{G}}$ in Figure 5.1 for the HB, NAG, and TM methods is given in Table 5.1. Note that for $\gamma = 0$ and $\gamma = \beta$, the TM controller reduces to the HB and NAG controllers, respectively. Moreover, for $\beta = \gamma = 0$, HB, NAG, and TM controllers reduce to the GD controller, with minimal realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = (\mathbf{1}, \alpha\mathbf{1}, \mathbf{1}, \mathbf{0})$.

Again, the controllers in Table 5.1 are strictly proper ($\mathbf{D} = \mathbf{0}$). Therefore, to apply the passivity-based analysis proposed in Section 4.3, the loop transformation in Figure 5.2 is performed. Notably, the resulting modified controller $\bar{\boldsymbol{\mathcal{G}}}$ has the same minimal realization as $\boldsymbol{\mathcal{G}}$, but with a non-zero $\mathbf{D} = D\mathbf{1}$. As such, the remainder of this chapter determines the lower bound on the feedthrough term D to render each modified controller $\bar{\boldsymbol{\mathcal{G}}}_{\text{HB}}$, $\bar{\boldsymbol{\mathcal{G}}}_{\text{NAG}}$, and $\bar{\boldsymbol{\mathcal{G}}}_{\text{TM}}$ passive.

5.3 Passivity of Modified Momentum-Based Controllers

In Lemma 4.1, it was shown that $\bar{\boldsymbol{\mathcal{G}}}_{\text{GD}}$ in Figure 4.4 is passive if and only if $D \geq \alpha/2$. This section provides similar lower bounds on the feedthrough term D for the HB, NAG, and TM controllers in Table 5.1.

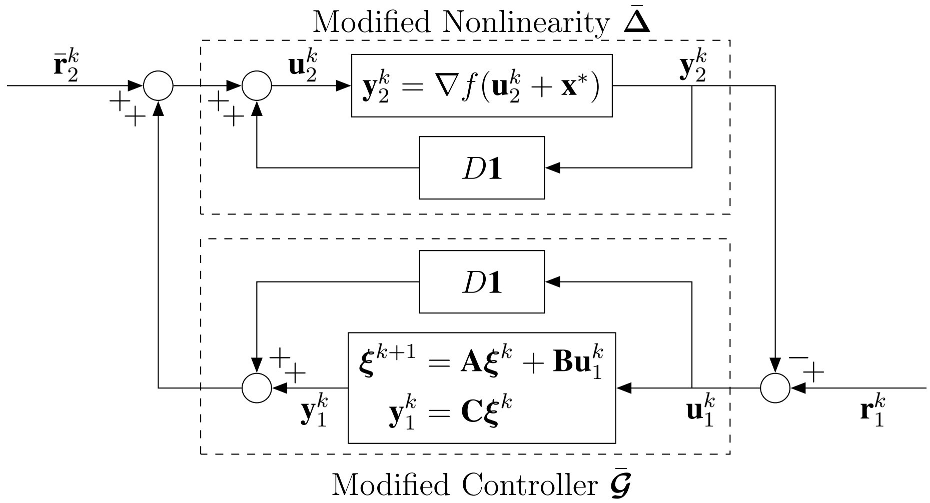


Figure 5.2: Loop transformation of the negative feedback interconnection in Figure 5.1. The loop transformation introduces a feedthrough term $D\mathbf{1}$, resulting in the modified controller $\bar{\boldsymbol{\mathcal{G}}}$, the modified nonlinearity $\bar{\mathbf{\Delta}}$, and $\bar{\mathbf{r}}_2^k = \mathbf{r}_2^k - D\mathbf{r}_1^k$.

5.3.1 Modified Heavy Ball Controller

For simplicity, consider the scalar case of $x \in \mathbb{R}$ such that the minimal realization of the modified HB controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{HB}}$ in Figure 4.4 is given by

$$\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{HB}} \overset{\min}{\sim} \left[\begin{array}{cc|c} 1+\beta & -\beta & \alpha \\ 1 & 0 & 0 \\ \hline 1 & 0 & D \end{array}\right], \tag{5.7}$$

for $\alpha, \beta, D \in \mathbb{R}_{>0}$. For a brief justification of this simplification, refer to Appendix B.

Lemma 5.1. The modified HB controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{HB}}$ in Figure 5.2 is passive if $D \geq \frac{\alpha(1+\beta)}{2(1-\beta)^2}$ and $\beta < 1$.

Proof. Following Lemma 4.1 and Remark 3.1, the modified HB controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{HB}}$ is passive if and only if

$$\exists \mathbf{P} = \begin{bmatrix} p_1 & p_2 \\ * & p_3 \end{bmatrix} \in \mathbb{S}^2_{++}, \tag{5.8}$$

such that

$$\mathbf{M} = \begin{bmatrix} p_3 - p_1 + 2p_2(\beta+1) + p_1(\beta+1)^2 & * & * \\ -(p_2 + \beta p_1)(\beta+1) & \beta^2 p_1 - p_3 & * \\ \alpha(p_2 + p_1(\beta+1)) - 1 & -\alpha\beta p_1 & \alpha^2 p_1 - 2D \end{bmatrix} \in \mathbb{S}^3_{-}. \tag{5.9}$$

Applying the principal minor test in Theorem 2.1, it follows that $\mathbf{P} \in \mathbb{S}^2_{++}$ if and only if

$$p_1 > 0, \qquad p_3 > 0, \qquad p_1 p_3 > p_2^2. \tag{5.10}$$

Similarly, $\mathbf{M} \in \mathbb{S}^3_{-}$ if and only if

$$\begin{aligned}
2p_2 + p_3 + 2\beta p_1 + 2\beta p_2 + \beta^2 p_1 &\leq 0, && (5.11a)\\
\beta^2 p_1 &\leq p_3, && (5.11b)\\
\alpha^2 p_1 &\leq 2D, && (5.11c)\\
p_2 + p_3 + \beta p_1 + \beta p_2 &= 0, && (5.11d)\\
\alpha^2(p_1^2 + p_2^2 - p_1 p_3) + 2D\big(2p_2 + p_3 + 2\beta(p_1 + p_2) + \beta^2 p_1\big) + 1 &\leq 2\alpha(p_2 + p_1 + \beta p_1), && (5.11e)\\
\alpha^2 p_1 p_3 &\leq 2D\big(p_3 - \beta^2 p_1\big), && (5.11f)
\end{aligned}$$

$$\begin{aligned}
&\alpha^2(p_1^2 p_3 - p_1 p_2^2 - p_1 p_3^2 + p_2^2 p_3) - 2\alpha(\beta p_1 p_2 + \beta p_1 p_3 + p_1 p_3 + p_2 p_3) - \beta^2 p_1 + p_3 \\
&+2D\big(\beta^2(p_1^2 + 2p_1 p_2 + p_2^2) + 2\beta(p_1 p_2 + p_1 p_3 + p_2^2 + p_2 p_3) + p_2^2 + 2p_2 p_3 + p_3^2\big) \leq 0.
\end{aligned} \tag{5.11g}$$

Note, since the goal is to establish a lower bound on D, the inequalities (5.11c) and (5.11f) are particularly useful. Firstly, from (5.11c), if $D = \alpha^2 p_1/2$, it can be shown that (5.11f) leads to a contradiction, since $p_1, \alpha \in \mathbb{R}_{>0}$. As such,

$$\alpha^2 p_1 < 2D. \tag{5.12}$$

Moreover, from (5.11b), if $p_3 = \beta^2 p_1$, again, it can be shown that (5.11f) leads to a contradiction. Therefore, it follows that

$$\beta^2 p_1 < p_3. \tag{5.13}$$

Using (5.13) and (5.11f), assume

$$D = \frac{\alpha^2 p_1 p_3}{2(p_3 - \beta^2 p_1)}. \tag{5.14}$$

Substituting (5.14) into (5.11g) and completing the square yields

$$\big(p_3 - \beta^2 p_1\big)\big(\beta^2 p_1 - p_3 + \alpha p_1 p_3 + \alpha p_2 p_3 + \alpha\beta p_1 p_2 + \alpha\beta p_1 p_3\big)^2 \leq 0. \tag{5.15}$$

From (5.13) and (5.15), it follows that

$$\beta^2 p_1 - p_3 + \alpha p_1 p_3 + \alpha p_2 p_3 + \alpha\beta p_1 p_2 + \alpha\beta p_1 p_3 = 0. \tag{5.16}$$

Solving for p_2 in (5.16) yields

$$p_2 = \frac{1}{\alpha} - p_1(\beta+1) + \frac{\beta p_1(\alpha p_1 - 1)(\beta+1)}{\alpha(p_3 + \beta p_1)}. \tag{5.17}$$

Substituting (5.17) into (5.11d) and rearranging yields

$$\frac{(p_3 - \beta^2 p_1)(\beta - \alpha p_1 + \alpha p_3 + 1)}{\alpha(p_3 + \beta p_1)} = 0. \tag{5.18}$$

From (5.13) and (5.18), it follows that

$$\beta - \alpha p_1 + \alpha p_3 + 1 = 0. \tag{5.19}$$

Solving for p_1 from (5.19) yields

$$p_1 = \frac{1+\beta}{\alpha} + p_3. \tag{5.20}$$

Substituting (5.20) into (5.13), it follows that

$$\beta^3 + \beta^2 < \alpha p_3\big(1 - \beta^2\big). \tag{5.21}$$

Assuming $\beta < 1$, from (5.21), it follows that

$$\frac{\beta^2}{\alpha(1-\beta)} < p_3. \tag{5.22}$$

Similarly, it can be shown that (5.12) and (5.13) are also satisfied, provided (5.22) holds. Moreover, substituting (5.17) and (5.20) into (5.11a) and collecting the p_3 terms yields the following quadratic inequality

$$c_1 p_3^2 + c_2 p_3 + c_3 \leq 0, \tag{5.23}$$

where

$$c_1 = -4\alpha^2(1-\beta)(\beta+1)^2, \qquad c_2 = 4\alpha\beta(2\beta-1)(\beta+1)^2, \qquad c_3 = 4\beta^3(\beta+1)^2. \tag{5.24}$$

The roots of (5.23) are given by

$$r_1 = -\frac{\beta}{\alpha}, \qquad\qquad r_2 = \frac{\beta^2}{\alpha(1-\beta)}. \tag{5.25}$$

Since $c_1 < 0$, and to satisfy $p_3 > 0$, it follows from (5.23) that $r_2 \leq p_3$. Therefore, the condition in (5.22) also ensures that (5.23) is satisfied. What remains is to find an expression for D in terms of α and β. Substituting (5.20) into (5.14) yields

$$D = \frac{\alpha^2 p_3(\beta + \alpha p_3 + 1)}{2(\beta+1)(\alpha p_3(1-\beta) - \beta^2)}. \tag{5.26}$$

It can be shown that for

$$p_3 = \frac{\beta(1+\beta)}{\alpha(1-\beta)}, \tag{5.27}$$

D in (5.26) is minimized and given by

$$D = \frac{\alpha(1+\beta)}{2(1-\beta)^2}. \tag{5.28}$$

As such, for

$$p_1 = \frac{1+\beta}{\alpha(1-\beta)}, \qquad p_2 = -\frac{2\beta}{\alpha(1-\beta)}, \qquad p_3 = \frac{\beta(1+\beta)}{\alpha(1-\beta)}, \qquad D = \frac{\alpha(1+\beta)}{2(1-\beta)^2}, \tag{5.29}$$

and $\beta < 1$, there exists a $\mathbf{P} \in \mathbb{S}^2_{++}$ such that $\mathbf{M} \in \mathbb{S}^3_{-}$. For completeness, the eigenvalues of $\mathbf{P}$ are given by

$$\lambda_{1,2}(\mathbf{P}) = \frac{1}{2\alpha(1-\beta)}\Big((1+\beta)^2 \pm \sqrt{\beta^4 + 14\beta^2 + 1}\Big). \tag{5.30}$$

Moreover, for $D = \frac{\alpha(1+\beta)}{2(1-\beta)^2}$, the eigenvalues of $\mathbf{M}$ are given by

$$\lambda_1(\mathbf{M}) = 0, \qquad \lambda_2(\mathbf{M}) = 0, \qquad \lambda_3(\mathbf{M}) = -\frac{\beta(1+\beta)\big(2(1-\beta)^2 + \alpha^2\big)}{\alpha(1-\beta)^2}. \tag{5.31}$$

□

5.3.2 Modified Nesterov's Accelerated Gradient Controller

For simplicity, consider the scalar case of $x \in \mathbb{R}$ such that the minimal realization of the modified NAG controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{NAG}}$ in Figure 4.4 is given by

$$\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{NAG}} \overset{\min}{\sim} \left[\begin{array}{cc|c} 1+\beta & -\beta & \alpha \\ 1 & 0 & 0 \\ \hline 1+\beta & -\beta & D \end{array}\right], \tag{5.32}$$

for $\alpha, \beta, D \in \mathbb{R}_{>0}$.

Lemma 5.2. The modified NAG controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{NAG}}$ in Figure 5.2 is passive if $D \geq \alpha + \frac{\alpha(3\beta-1)}{2(1-\beta)^2}$ and $\beta < 1$.

Proof. Similar to the proof of Lemma 5.1, it can be shown that for

$$p_1 = \frac{1+\beta^2}{\alpha(1-\beta)}, \qquad p_2 = -\frac{\beta(1+\beta)}{\alpha(1-\beta)}, \qquad p_3 = \frac{2\beta^2}{\alpha(1-\beta)}, \qquad D = \alpha + \frac{\alpha(3\beta-1)}{2(1-\beta)^2}, \tag{5.33}$$

and $\beta < 1$, there exists a $\mathbf{P} \in \mathbb{S}^2_{++}$ such that $\mathbf{M} \in \mathbb{S}^3_{-}$. Again, for completeness, the eigenvalues of $\mathbf{P}$ are given by

$$\lambda_{1,2}(\mathbf{P}) = \frac{1}{2\alpha(1-\beta)}\Big(3\beta^2 + 1 \pm (1+\beta)\sqrt{5\beta^2 - 2\beta + 1}\Big). \tag{5.34}$$

Moreover, for $D = \alpha + \frac{\alpha(3\beta-1)}{2(1-\beta)^2}$, the eigenvalues of $\mathbf{M}$ are given by

$$\lambda_1(\mathbf{M}) = 0, \qquad \lambda_2(\mathbf{M}) = 0, \qquad \lambda_3(\mathbf{M}) = -\frac{\beta^2(1+\beta)\big(2(1-\beta)^2 + \alpha^2\big)}{\alpha(1-\beta)^2}. \tag{5.35}$$

□

5.3.3 Modified Triple Momentum Controller

Again, consider the scalar case of $x \in \mathbb{R}$ such that the minimal realization of the modified TM controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{TM}}$ in Figure 4.4 is given by

$$\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{TM}} \overset{\min}{\sim} \left[\begin{array}{cc|c} 1+\beta & -\beta & \alpha \\ 1 & 0 & 0 \\ \hline 1+\gamma & -\gamma & D \end{array}\right], \tag{5.36}$$

for $\alpha, \beta, \gamma, D \in \mathbb{R}_{>0}$.

Lemma 5.3. The modified TM controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{TM}}$ in Figure 5.2 is passive if

$$D \geq \begin{cases} \dfrac{\alpha(2\gamma+1)}{2(\beta+1)}, & \beta < \dfrac{\gamma}{\gamma+1}, \\ \dfrac{\alpha}{(1-\beta)^2} - \dfrac{\alpha(2\gamma+1)}{2(1-\beta)}, & \dfrac{\gamma}{\gamma+1} < \beta < 1. \end{cases} \tag{5.37}$$

Proof. Following Lemma 4.1 and Remark 3.1, the modified NAG controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{NAG}}$ is passive if and only if

$$\exists \mathbf{P} = \begin{bmatrix} p_1 & p_2 \\ * & p_3 \end{bmatrix} \in \mathbb{S}^2_{++}, \tag{5.38}$$

such that

$$\mathbf{M} = \begin{bmatrix} p_3 - p_1 + 2p_2(\beta+1) + p_1(\beta+1)^2 & * & * \\ -(p_2+\beta p_1)(\beta+1) & \beta^2 p_1 - p_3 & * \\ \alpha(p_2 + p_1(\beta+1)) - \gamma - 1 & \gamma - \alpha\beta p_1 & \alpha^2 p_1 - 2D \end{bmatrix} \in \mathbb{S}^3_-. \tag{5.39}$$

Applying the principal minor test in Theorem 2.1, it follows that $\mathbf{P} \in \mathbb{S}^2_{++}$ if and only if

$$p_1 > 0, \qquad p_3 > 0, \qquad p_1 p_3 > p_2^2. \tag{5.40}$$

Similarly, $\mathbf{M} \in \mathbb{S}^3_-$ if and only if

$$2p_2 + p_3 + 2\beta p_1 + 2\beta p_2 + \beta^2 p_1 \leq 0, \tag{5.41a}$$

$$\beta^2 p_1 \leq p_3, \tag{5.41b}$$

$$\alpha^2 p_1 \leq 2D, \tag{5.41c}$$

$$p_2 + p_3 + \beta p_1 + \beta p_2 = 0, \tag{5.41d}$$

$$\begin{aligned} &\alpha^2(p_1^2 + p_2^2 - p_1p_3) + 2D\big(2p_2 + p_3 + 2\beta(p_1+p_2) + \beta^2 p_1\big) + 1 \\ &\qquad -\gamma(2\alpha(p_1 + \beta p_1 + p_2) - \gamma - 2) \leq 2\alpha(p_2 + p_1 + \beta p_1), \end{aligned} \tag{5.41e}$$

$$\gamma^2 - 2\alpha\beta\gamma p_1 + \alpha^2 p_1 p_3 \leq 2D\big(p_3 - \beta^2 p_1\big), \tag{5.41f}$$

$$\begin{aligned} &\alpha^2(p_1^2p_3 - p_1p_2^2 - p_1p_3^2 + p_2^2p_3) - \beta^2 p_1 + p_3 + 2\gamma(p_2 + p_3 + \beta(p_1+p_2)) \\ &-2\alpha\big(\beta p_1p_2 + \beta p_1p_3 + p_1p_3 + p_2p_3 + \gamma\big(p_2^2 + p_1p_2 + p_1p_3 + p_2p_3 + \beta(p_1+p_2)^2\big)\big) \\ &+2D\big(\beta^2(p_1^2 + 2p_1p_2 + p_2^2) + 2\beta(p_1p_2 + p_1p_3 + p_2^2 + p_2p_3) + p_2^2 + 2p_2p_3 + p_3^2\big) \leq 0. \end{aligned} \tag{5.41g}$$

From $p_1p_3 > p_2^2$, (5.41b) and (5.41d), it can be shown that $\beta^2 p_1 = p_3$ leads to a contradiction. As such, it follows that

$$\beta^2 p_1 < p_3. \tag{5.42}$$

Here, contrary to Lemma 5.1 and Lemma 5.2, the equality of (5.41c) does not lead to a contradiction. Consider the case where $D = \alpha^2 p_1/2$. Assuming $\beta < \frac{\gamma}{1+\gamma}$, it can be shown that for

$$p_1 = \frac{\gamma}{\alpha\beta}, \qquad p_2 = \frac{\beta-\gamma}{\alpha\beta}, \qquad p_3 = \frac{\gamma - \beta - \beta^2}{\alpha\beta}, \qquad D = \frac{\alpha\gamma}{2\beta}, \tag{5.43}$$

there exists a $\mathbf{P} \in \mathbb{S}^2_{++}$ such that $\mathbf{M} \in \mathbb{S}^3_-$.

Now, from (5.41f), consider the case where $\gamma^2 - 2\alpha\beta\gamma p_1 + \alpha^2 p_1 p_3 = 2D(p_3 - \beta^2 p_1)$. It follows that

$$\frac{\gamma^2 - 2\alpha\beta\gamma p_1 + \alpha^2 p_1 p_3}{2(p_3 - \beta^2 p_1)} = D. \tag{5.44}$$

Moreover, from (5.41d), it follows that

$$p_2 = -\frac{p_3 + \beta p_1}{1+\beta}. \tag{5.45}$$

Substituting (5.44) and (5.45) into (5.41g) and completing the square yields

$$\left(p_3 - \beta^2 p_1\right)\left(\beta - \alpha p_1 + \alpha p_3 + 1\right)^2 \le 0. \tag{5.46}$$

As per (5.42), $p_3 - \beta^2 p_1 > 0$. Consequently, (5.46) implies that

$$\beta - \alpha p_1 + \alpha p_3 + 1 = 0. \tag{5.47}$$

Rearranging (5.47) yields

$$p_1 = \frac{1+\beta}{\alpha} + p_3. \tag{5.48}$$

Assuming $\beta < 1$, substituting (5.48) into (5.42), it follows that

$$\frac{\beta^2}{\alpha(1-\beta)} < p_3. \tag{5.49}$$

Moreover, substituting (5.48) into (5.44) yields

$$D = \frac{-\alpha^3 p_3^2 + (2\beta\gamma - 1 - \beta)\alpha^2 p_3 + (2\beta^2 + 2\beta - \gamma)\alpha\gamma}{2((\beta^2 - 1)\alpha p_3 + \beta^3 + \beta^2)}. \tag{5.50}$$

What remains is to find an expression for p_3 such that D is minimized. To this end, the analysis is split into two cases based on the value of β.

Case 1: $\beta < \frac{\gamma}{\gamma+1}$.
It can be shown that for

$$p_3 = \frac{\gamma - \beta}{\alpha}, \tag{5.51}$$

D in (5.50) is minimized and given by

$$D = \frac{\alpha(2\gamma+1)}{2(\beta+1)}. \tag{5.52}$$

Notably, for $\beta < \frac{\gamma}{\gamma+1}$, the D in (5.50) is strictly smaller than that of (5.43). As such, if $\beta < \frac{\gamma}{1+\gamma}$, for

$$p_1 = \frac{1+\gamma}{\alpha}, \qquad p_2 = -\frac{\gamma}{\alpha}, \qquad p_3 = \frac{\gamma-\beta}{\alpha}, \qquad D = \frac{\alpha(2\gamma+1)}{2(\beta+1)}, \tag{5.53}$$

there exists a $\mathbf{P} \in \mathbb{S}^2_{++}$ such that $\mathbf{M} \in \mathbb{S}^3_{-}$. For $\beta < \frac{\gamma}{1+\gamma}$ and p_1, p_2, and p_3 as in (5.53), the eigenvalues of $\mathbf{P}$ are given by

$$\lambda_{1,2}(\mathbf{P}) = \frac{1}{2\alpha}\Big(2\gamma - \beta + 1 \pm \sqrt{\beta^2 + 2\beta + 4\gamma^2 + 1}\Big). \tag{5.54}$$

Moreover, for $D = \frac{\alpha(2\gamma+1)}{2(\beta+1)}$, the eigenvalues of $\mathbf{M}$ are given by

$$\lambda_1(\mathbf{M}) = 0, \qquad \lambda_2(\mathbf{M}) = 0, \qquad \lambda_3(\mathbf{M}) = \frac{(\beta - \gamma + \beta\gamma)(\alpha^2 + 2(1+\beta)^2)}{\alpha(\beta+1)}. \tag{5.55}$$

Case 2: $\frac{\gamma}{\gamma+1} < \beta < 1$.
Similar to Case 1, it can be shown that for

$$p_3 = \frac{\beta - \gamma + \beta\gamma + \beta^2}{\alpha(1-\beta)}, \tag{5.56}$$

D in (5.50) is minimized and given by

$$D = \frac{\alpha}{(1-\beta)^2} - \frac{\alpha(2\gamma+1)}{2(1-\beta)}. \tag{5.57}$$

As such, if $\frac{\gamma}{\gamma+1} < \beta < 1$, for

$$\begin{gathered} p_1 = \frac{2}{\alpha(1-\beta)} - \frac{1+\gamma}{\alpha}, \qquad p_2 = \frac{2+\gamma}{\alpha} - \frac{2}{\alpha(1-\beta)}, \qquad p_3 = \frac{\beta - \gamma + \beta\gamma + \beta^2}{\alpha(1-\beta)}, \\ D = \frac{\alpha}{(1-\beta)^2} - \frac{\alpha(2\gamma+1)}{2(1-\beta)}, \end{gathered} \tag{5.58}$$

there exists a $\mathbf{P} \in \mathbb{S}^2_{++}$ such that $\mathbf{M} \in \mathbb{S}^3_{-}$. For $\beta < \frac{\gamma}{1+\gamma}$ and p_1, p_2, and p_3 as in (5.53), the eigenvalues of $\mathbf{P}$ are given by

$$\lambda_{1,2}(\mathbf{P}) = \frac{2(\beta - \gamma + \beta\gamma) + \beta^2 + 1 \pm \sqrt{\beta^4 + 4\beta^2\gamma^2 + 16\beta^2\gamma + 14\beta^2 - 8\beta\gamma^2 - 16\beta\gamma + 4\gamma^2 + 1}}{2\alpha(1-\beta)}. \tag{5.59}$$

Moreover, for $D = \frac{\alpha}{(1-\beta)^2} - \frac{\alpha(2\gamma+1)}{2(1-\beta)}$, the eigenvalues of $\mathbf{M}$ are given by

$$\lambda_1(\mathbf{M}) = 0, \quad \lambda_2(\mathbf{M}) = 0, \quad \lambda_3(\mathbf{M}) = -\frac{(\beta+1)(\beta - \gamma + \beta\gamma)(\alpha^2 + 2(1-\beta)^2)}{\alpha(1-\beta)^2}. \tag{5.60}$$

□

Table 5.2: Lower Bound on the Feedthrough Term D Needed to Render the Modified First-Order Optimization Controllers Passive.

$\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GD}}$	$\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{HB}}$	$\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{NAG}}$	$\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{TM}}$
$D = \frac{\alpha}{2}$	$D = \frac{\alpha(1+\beta)}{2(1-\beta)^2}$	$D = \alpha + \frac{\alpha(3\beta-1)}{2(1-\beta)^2}$	$D = \begin{cases} \frac{\alpha(2\gamma+1)}{2(\beta+1)}, & \beta < \frac{\gamma}{\gamma+1}, \\ \frac{\alpha}{(1-\beta)^2} - \frac{\alpha(2\gamma+1)}{2(1-\beta)}, & \frac{\gamma}{\gamma+1} < \beta. \end{cases}$

5.4 Discussion

In Section 5.3, the lower bound on the feedthrough term D needed to render the modified momentum-based controllers passive were derived. The results of Lemma 4.1, Lemma 5.1, Lemma 5.2, and Lemma 5.3 are summarized in Table 5.2. Notably, for $\beta = \gamma = 0$, all modified momentum-based controllers reduce to the modified GD controller, with a lower bound of $D = \alpha/2$. Moreover, for $\gamma = 0$, the lower bound on D for the modified TM controller reduces to that of the modified HB controller. Finally, for $\gamma = \beta$, the lower bound on D for the modified TM controller reduces to that of the modified NAG controller.

5.4.1 Input-Output Stability of Momentum-Based Methods

As discussed in Section 4.4, the advantage of this proposed passivity-based analysis is that it decouples the stability analysis of the controller from the properties of the objective function. As such, for a function $f \in \mathcal{S}_{0,L}$ with $0 < L$, the results of Lemma 4.5 can be reused to guarantee the modified nonlinearity $\bar{\boldsymbol{\Delta}}$ is OSP, provided $D < 1/L$. What follows is the extension of Theorem 4.3 to the modified momentum-based controllers in Table 5.1.

Theorem 5.1. Consider the negative feedback interconnection in Figure 5.2, where $f \in \mathcal{S}_{0,L}$, $\mathbf{r}_1 = \mathbf{0}$, and $\bar{\mathbf{r}}_2 = \mathbf{0}$. The following special cases hold:

(i) For $\bar{\boldsymbol{\mathcal{G}}} = \bar{\boldsymbol{\mathcal{G}}}_{\mathrm{HB}}$, if

$$\frac{\alpha(1+\beta)}{2(1-\beta)^2} < \frac{1}{L}, \tag{5.61}$$

and $\beta < 1$, then $\nabla f(\mathbf{x}^k) \to \mathbf{0}$.

(ii) For $\bar{\boldsymbol{\mathcal{G}}} = \bar{\boldsymbol{\mathcal{G}}}_{\mathrm{NE}}$, if

$$\alpha + \frac{\alpha(3\beta-1)}{2(1-\beta)^2} < \frac{1}{L}, \tag{5.62}$$

and $\beta < 1$, then $\nabla f(\mathbf{x}^k) \to \mathbf{0}$.

Table 5.3: Optimal Tuning Parameters of Momentum-Based Methods for $f \in \mathcal{F}_{m,L}$.

	Optimal Tuning Parameters		
HB	$\alpha = \frac{4}{\left(\sqrt{L}+\sqrt{m}\right)^2}$	$\beta = \frac{\left(\sqrt{L}-\sqrt{m}\right)^2}{\left(\sqrt{L}+\sqrt{m}\right)^2}$	
NAG	$\alpha = \frac{1}{L}$	$\beta = \frac{\sqrt{L}-\sqrt{m}}{\sqrt{L}+\sqrt{m}}$	
TM	$\alpha = \frac{2}{L} - \sqrt{\frac{m}{L^3}}$	$\beta = \frac{(\sqrt{L}-\sqrt{m})^3}{\sqrt{L}(L-m)}$	$\gamma = \frac{5L-m}{2L-m+\sqrt{Lm}} - 2$

(iii) For $\bar{\boldsymbol{\mathcal{G}}} = \bar{\boldsymbol{\mathcal{G}}}_{\mathrm{TM}}$, if

$$\begin{cases} \dfrac{\alpha(2\gamma+1)}{2(\beta+1)} < \dfrac{1}{L}, & \beta < \dfrac{\gamma}{\gamma+1}, \\ \dfrac{\alpha}{(1-\beta)^2} - \dfrac{\alpha(2\gamma+1)}{2(1-\beta)} < \dfrac{1}{L}, & \dfrac{\gamma}{\gamma+1} < \beta < 1, \end{cases} \tag{5.63}$$

then $\nabla f(\mathbf{x}^k) \to \mathbf{0}$.

Proof. As per Lemma 4.5, provided $D < 1/L$, $\bar{\boldsymbol{\Delta}}$ is OSP. Moreover, Lemma 5.1, Lemma 5.2, and Lemma 5.3 provide the lower bounds on D needed to render $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{HB}}$, $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{NAG}}$, and $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{TM}}$ passive, respectively. It follows from Theorem 3.1-(ii) t hat $\mathbf{y}_2 \in \ell_2$. Therefore, $\mathbf{y}_2^k = \nabla f(\boldsymbol{\xi}^k + \mathbf{x}^*) = \nabla f(\mathbf{x}^k) \to \mathbf{0}$. □

5.4.2 Comparing with Existing Results

The results of Theorem 5.1 provide sufficient conditions for the global convergence of the modified HB, NAG, and TM methods when $f \in \mathcal{S}_{0,L}$. In comparison, [8] states that the global convergence of the HB method is guaranteed for $f \in \mathcal{S}_{m,L}$ with $0 < m < L$ provided that

$$\alpha < \begin{cases} \dfrac{2(1+\beta)}{L}, & \beta \leq \bar{\beta}, \\ \dfrac{2(1-\beta)^2}{(1+\beta)(L+m) - 4\sqrt{\beta L m}}, & \beta > \bar{\beta}, \end{cases} \tag{5.64}$$

where $\bar{\beta} = \left(\sqrt{\frac{L}{m}} - \sqrt{\frac{L}{m} - 1}\right)^2$. Notice that as $m \to 0$, $\bar{\beta} \to 0$, and (5.64) reduces to (5.61).

As shown in [6, 7, 15], for $f \in \mathcal{F}_{m,L}$ with $0 < m \leq L$, the optimal tuning parameters for HB, NAG, and TM are summarized in Table 5.3. Since $\mathcal{F}_{m,L} \subset \mathcal{S}_{0,L}$, it is of interest to determine for what range of condition numbers, $\kappa = L/m$, can Theorem 5.1 guarantee the global convergence of the HB, NAG, and TM methods using the optimal tuning parameters in Table 5.3.

In the case of HB, substituting the optimal tuning parameters into (5.61), it follows that Theorem 5.1-(i) can only guarantee the global convergence of the HB method if

$$\frac{L}{m} = \kappa_{\mathrm{HB}} < 3. \tag{5.65}$$

The bound in (5.65) is more conservative than that of [8, Theorem 1], which guarantees the global convergence of the HB method for $\kappa_{\mathrm{HB}} < 3 + 2\sqrt{2}$, for $f \in \mathcal{S}_{m,L}$. Similarly, for NAG, substituting the optimal tuning parameters into (5.62), it follows that Theorem 5.1-(ii) can only guarantee the global convergence of the NAG method if

$$\frac{L}{m} = \kappa_{\mathrm{NAG}} < 4. \tag{5.66}$$

Finally, for TM, the optimal tuning parameters in Table 5.3 satisfy the case of $\frac{\gamma}{1+\gamma} < \beta$. As such, it follows that Theorem 5.1-(iii) can only guarantee the global convergence of the TM method if

$$\frac{L}{m} = \kappa_{\mathrm{TM}} < 3.66985. \tag{5.67}$$

Again, the bound in (5.67) is more conservative than that of [8, Theorem 2], which guarantees the global convergence of the TM method for $\kappa_{\mathrm{TM}} < 8.1776$, for $f \in \mathcal{S}_{m,L}$.

5.5 Closing Remarks

This chapter presented the extension of the passivity-based analysis proposed in Chapter 4 to popular first-order momentum-based methods such as HB, NAG, and TM. Sufficient conditions for the passivity of the modified momentum-based controllers were analytically derived. Notably, for $\beta = \gamma = 0$, all modified momentum-based controllers reduced to the modified GD controller, with a lower bound $D = \alpha/2$. Moreover, for $\gamma = 0$, this bound for the modified TM controller reduced to that of the modified HB controller. Similarly, for $\gamma = \beta$, the modified TM controller reduced to that of the modified NAG controller. For $f \in \mathcal{S}_{0,L}$, the proposed passivity-based analysis guarantees the global convergence of the momentum-based methods for a broader class of functions than previously considered in the literature. Finally, by taking $m \to 0$, the existing bounds for the global convergence of the HB method for $f \in \mathcal{S}_{m,L}$ with $0 < m < L$ were shown to recover the proposed bounds for the modified HB method for $f \in \mathcal{S}_{0,L}$ with $0 < L$.

Part 2

Gain-Scheduling of QSR-Dissipative Systems

Chapter 6

Matrix-Scheduling of QSR-Dissipative Systems

6.1 Overview

This chapter presents the gain-scheduling of QSR-dissipative systems and first-order optimization algorithms.

First, this chapter presents a generalized matrix-gain-scheduling architecture along with a detailed discussion on the *design* and *construction* of scheduling matrices. The gain-scheduling results in [21] and [3] are extended by considering the scheduling of a broader class of QSR-dissipative systems using the proposed matrix-gain-scheduling architecture. The scheduling of important special cases of QSR-dissipativity, such as passive, ISP, OSP, finite $\mathcal{L}_2$ gain, VSP, and conic systems are shown to result in an overall gain-scheduled system of the same special case.

Second, leveraging the passivity-based analysis of first-order optimization algorithms presented in Part 1, a gain-scheduled interpretation is given for first-order optimization algorithms with varying hyperparameters. Moreover, a new variation of the gradient descent algorithm is proposed, in which both the input and output of the gradient are scaled. Finally, the gain-scheduling of passive first-order optimization algorithms is considered, and it is shown that, provided the scheduling functions satisfy a certain upper bound, the global convergence of the gain-scheduled algorithm is guaranteed for $f \in \mathcal{S}_{0,L}$ with $0 < L$.

6.2 Preliminaries

This section presents the discrete-time definition of QSR-dissipative systems, where the truncated inner product is defined over the discrete-time interval $\mathcal{T} = \{0, 1, \ldots, T-1\}$ for $T \in \mathbb{Z}_{>0}$.

6.2.1 Characterization of QSR-Dissipativity

Dissipative systems theory [31] is a general and broadly applicable systems theoretic framework that characterize a dynamical system by its input-output behavior, and it strongly relates to Lyapunov and $\mathcal{L}_2$ stability theories [39]. As discussed in [40], the input-output dissipativity theory in [5, 30] is an extension of classical Lyapunov based dissipativity theory in [31, 41], where *a priori* information between input and output variables is not distinguished. Through the introduction of a storage function, these two approaches were unified in [42]. One such class of storage functions are quadratic supply rate (QSR) functions, which are used to analyze the stability of interconnected systems [41, 43]. A special case of QSR-dissipativity for square systems, whose inputs and outputs have the same dimension, is passivity. Over the years, passivity and dissipativity have seen a wide range of applications, including exponential stabilization of dynamical systems [44], analysis of event-triggered networked control systems [45], synchronization of neural networks [46], and control of Euler–Lagrange systems [47], flexible multi-link manipulators [48], and delayed teleoperations [49]. The following definition characterizes QSR-dissipativity of a discrete-time system.

Definition 6.1 (QSR-Dissipativity [21, 31, 41])**.** Consider a causal discrete-time system, $\boldsymbol{\mathcal{G}} : \ell_{2e} \to \ell_{2e}$, defined by

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \qquad \mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t)), \tag{6.1}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$, $\mathbf{y} \in \mathbb{R}^{n_y}$, and $\mathbf{u} \in \mathbb{R}^{n_u}$. The system $\boldsymbol{\mathcal{G}}$ is QSR-dissipative with matrices $\mathbf{Q} \in \mathbb{S}^{n_y}$, $\mathbf{S} \in \mathbb{R}^{n_y \times n_u}$, and $\mathbf{R} \in \mathbb{S}^{n_u}$ if for all $\mathbf{u} \in \ell_{2e}$ and $T \in \mathbb{Z}_{>0}$ there exists a storage function $V : \mathbb{R}^{n_x} \to \mathbb{R}_{\geq 0}$ such that

$$\langle \mathbf{y}, \mathbf{Q}\mathbf{y} \rangle_T + 2 \langle \mathbf{y}, \mathbf{S}\mathbf{u} \rangle_T + \langle \mathbf{u}, \mathbf{R}\mathbf{u} \rangle_T \geq V(\mathbf{x}(T)) - V(\mathbf{x}(0)). \tag{6.2}$$

The special cases of QSR-dissipativity are

- *passive* if $\mathbf{Q} = \mathbf{0}$, $\mathbf{S} = \frac{1}{2}\mathbf{1}$, and $\mathbf{R} = \mathbf{0}$,
- *ISP* if $\mathbf{Q} = \mathbf{0}$, $\mathbf{S} = \frac{1}{2}\mathbf{1}$, and $\mathbf{R} = -\delta\mathbf{1}$, where $\delta \in \mathbb{R}_{>0}$,
- *OSP* if $\mathbf{Q} = -\varepsilon\mathbf{1}$, $\mathbf{S} = \frac{1}{2}\mathbf{1}$, and $\mathbf{R} = \mathbf{0}$, where $\varepsilon \in \mathbb{R}_{>0}$,
- *finite ℓ_2 gain* if $\mathbf{Q} = -\mathbf{1}$, $\mathbf{S} = \mathbf{0}$, and $\mathbf{R} = \gamma^2\mathbf{1}$, where $\gamma \in \mathbb{R}_{>0}$,
- *VSP* if $\mathbf{Q} = -\varepsilon\mathbf{1}$, $\mathbf{S} = \frac{1}{2}\mathbf{1}$, and $\mathbf{R} = -\delta\mathbf{1}$, where $\delta, \varepsilon \in \mathbb{R}_{>0}$, and
- *conic* if $\mathbf{Q} = -\mathbf{1}$, $\mathbf{S} = \frac{a+b}{2}\mathbf{1} = c\mathbf{1}$, and $\mathbf{R} = -ab\mathbf{1} = (r^2 - c^2)\mathbf{1}$, where $a, b \in \mathbb{R}$ represent lower and upper conic bounds, while $c \in \mathbb{R}$ and $r \in \mathbb{R}_{>0}$ represent the center and radius of the conic sector, respectively.

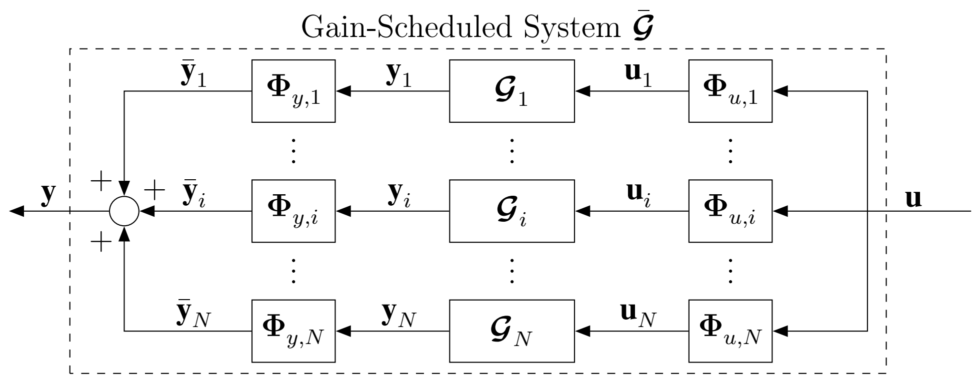


Figure 6.1: Gain-scheduled system $\bar{\boldsymbol{\mathcal{G}}}$, composed of N parallel QSR-dissipative subsystems. The input and output of each subsystem are scheduled as per (6.3) via the matrix multiplication between the scheduling matrices $\boldsymbol{\Phi}_{u,i}(t)$ and $\boldsymbol{\Phi}_{y,i}(t)$, and their corresponding signals $\mathbf{u}(t)$ and $\mathbf{y}_i(t)$, respectively.

6.3 Matrix-Gain-Scheduling Architecture

This section considers the matrix-gain-scheduling architecture in Figure 6.1, where the scheduling functions to be designed take the form of scheduling matrices, allowing for greater design freedom compared to the scalar scheduling signals in [2, 17–22, 50].

6.3.1 Input-Output Mapping

Consider the discrete-time gain-scheduled system, $\bar{\boldsymbol{\mathcal{G}}}$, in Figure 6.1, composed of N parallel QSR-dissipative subsystems, $\boldsymbol{\mathcal{G}}_1, \ldots, \boldsymbol{\mathcal{G}}_N$. The subsystems could be linear or nonlinear, and their input-output maps are given by $\mathbf{y}_i(t) = \boldsymbol{\mathcal{G}}_i(\mathbf{u}_i(t))$ for $i \in \mathcal{N} = \{1, \ldots, N\}$. The input and output of each subsystem are scheduled in a multiplicative manner such that

$$\mathbf{u}_i(t) = \boldsymbol{\Phi}_{u,i}(t)\mathbf{u}(t), \tag{6.3a}$$

$$\bar{\mathbf{y}}_i(t) = \boldsymbol{\Phi}_{y,i}(t)\mathbf{y}_i(t), \tag{6.3b}$$

where $\boldsymbol{\Phi}_{u,i} \in \mathbb{R}^{n_u \times n_u}$ and $\boldsymbol{\Phi}_{y,i} \in \mathbb{R}^{n_y \times n_y}$ are the scheduling matrices, $\mathbf{u}_i, \mathbf{u} \in \mathbb{R}^{n_u}$, and $\mathbf{y}_i, \mathbf{y} \in \mathbb{R}^{n_y}$, for all $i \in \mathcal{N}$. The discrete-time gain-scheduled system's input-output map can be written in terms of the individual subsystems inputs, outputs, and scheduling matrices as

$$\mathbf{y}(t) = \sum_{i \in \mathcal{N}} \bar{\mathbf{y}}_i(t) = \sum_{i \in \mathcal{N}} \boldsymbol{\Phi}_{y,i}(t)\mathbf{y}_i(t) = \sum_{i \in \mathcal{N}} \boldsymbol{\Phi}_{y,i}(t)\boldsymbol{\mathcal{G}}_i(\mathbf{u}_i(t)) = \sum_{i \in \mathcal{N}} \boldsymbol{\Phi}_{y,i}(t)\boldsymbol{\mathcal{G}}_i\big(\boldsymbol{\Phi}_{u,i}(t)\mathbf{u}(t)\big). \tag{6.4}$$

Based on the QSR-dissipativity property in Definition 6.1, the input and output of each subsystem must satisfy

$$\langle \mathbf{y}_i, \mathbf{Q}_i \mathbf{y}_i \rangle_T + 2 \langle \mathbf{y}_i, \mathbf{S}_i \mathbf{u}_i \rangle_T + \langle \mathbf{u}_i, \mathbf{R}_i \mathbf{u}_i \rangle_T \geq \tilde{V}_i, \qquad \forall T \in \mathbb{Z}_{>0}, \tag{6.5}$$

with $\tilde{V}_i = V_i(\mathbf{x}(T)) - V_i(\mathbf{x}(0))$, $\mathbf{Q} \in \mathbb{S}^{n_y}$, $\mathbf{S} \in \mathbb{R}^{n_y \times n_u}$, and $\mathbf{R} \in \mathbb{S}^{n_u}$ for all $i \in \mathcal{N}$ and $\mathbf{u}_i \in \ell_{2e}$.

The objective is to *design* the scheduling matrices $\mathbf{\Phi}_{u,i}(t)$ and $\mathbf{\Phi}_{y,i}(t)$ for $i \in \mathcal{N}$, such that the gain-scheduling of the N QSR-dissipative subsystems results in an overall gain-scheduled QSR-dissipative system. Furthermore, if all subsystems belong to the same special case of QSR-dissipativity as defined in Definition 6.1, the resulting overall gain-scheduled system must also belong to that same special case.

6.3.2 Scheduling Matrix Properties

Consider the set of scheduling matrices $\mathbf{\Phi}_{j,i}(t) \in \mathbb{R}^{n_j \times n_j}$ for $j \in \{u, y\}$ and $i \in \mathcal{N}$. With abuse of set notation, denote the time dependent set, $\mathcal{I}_j(t)$, as the index set of all the respective full rank scheduling matrices at time $t \in \mathbb{Z}_{\geq 0}$. That is, for $t \in \mathbb{Z}_{\geq 0}$ and $j \in \{u, y\}$,

$$\mathcal{I}_j(t) = \left\{ i \in \mathcal{N} \mid \operatorname{rank}\big(\mathbf{\Phi}_{j,i}(t)\big) = n_j \right\}. \tag{6.6}$$

Definition 6.2 (Active Scheduling Matrices)**.** Given a discrete-time gain-scheduled system of the type shown in Figure 6.1 with scheduling matrices $\mathbf{\Phi}_{j,i}(t) \in \mathbb{R}^{n_j \times n_j}$ for $j \in \{u, y\}$ and $i \in \mathcal{N}$, the scheduling matrices are said to be

- *active* if at all times, there exists at least one nonzero scheduling matrix, meaning $\forall t \in \mathbb{Z}_{\geq 0}$, $\exists i \in \mathcal{N}$ such that $\mathbf{\Phi}_{j,i}(t) \neq \mathbf{0}$ and
- *strongly active* if at all times, there exists at least one full rank scheduling matrix, meaning $\forall t \in \mathbb{Z}_{\geq 0}$, $\exists i \in \mathcal{N}$ such that $\operatorname{rank}\big(\mathbf{\Phi}_{j,i}(t)\big) = n_j$. Using set notation, this can be written as $\forall t \in \mathbb{Z}_{\geq 0}$, $\mathcal{I}_j(t) \neq \varnothing$.

When considering the special cases of QSR-dissipativity in Section 6.5, the active and strongly active cases of scheduling matrices are invoked to correctly quantify the overall QSR-dissipativity of the discrete-time gain-scheduled system.

For the remainder of this chapter, it is assumed that the scheduling matrices are designed such that

$$\sup_{t \in \mathbb{Z}_{\geq 0}} \left\| \mathbf{\Phi}_{j,i}(t) \right\|_2 = \sup_{t \in \mathbb{Z}_{\geq 0}} \sigma_{j,i}(t) = \left\| \sigma_{j,i} \right\|_\infty < \infty, \tag{6.7}$$

for all $j \in \{u, y\}$ and $i \in \mathcal{N}$, where $\sigma_{j,i}(t)$ is the largest singular value of $\mathbf{\Phi}_{j,i}(t)$. This can be thought of as an extension of the ℓ_∞ space in Definition 2.6 for matrix functions $\mathbf{\Phi}_{j,i} : \mathbb{Z}_{\geq 0} \to \mathbb{R}^{n_j \times n_j}$. Finally, for simplicity, the scheduling matrices are assumed to be explicitly time-dependent, though a similar analysis can be used to account for dependence on other signals or parameters.

6.3.3 Scheduling Matrix Construction

This subsection presents the construction of the scheduling matrices such that they satisfy the following pseudo-commutativity condition.

Definition 6.3 (Pseudo-Commuting Scheduling Matrices)**.** Consider a discrete-time gain-scheduled system of type shown in Figure 6.1 with scheduling matrices $\mathbf{\Phi}_{u,i}(t) \in \mathbb{R}^{n_u \times n_u}$ and $\mathbf{\Phi}_{y,i}(t) \in \mathbb{R}^{n_y \times n_y}$ for $i \in \mathcal{N}$. Given a series of matrices $\mathbf{S}_i \in \mathbb{R}^{n_y \times n_u}$ for $i \in \mathcal{N}$, the scheduling matrices are said to *pseudo-commute with* $\mathbf{S}_i$, if they satisfy

$$\mathbf{\Phi}_{y,i}^{\mathsf{T}}(t)\mathbf{S}_i = \mathbf{S}_i \mathbf{\Phi}_{u,i}(t), \qquad \forall t \in \mathbb{Z}_{\geq 0}, \forall i \in \mathcal{N}. \tag{6.8}$$

In Section 6.4, for each QSR-dissipative subsystem $\boldsymbol{\mathcal{G}}_i$ satisfying (6.5) with $\mathbf{Q}_i$, $\mathbf{S}_i$, and $\mathbf{R}_i$, this pseudo-commutativity condition is invoked by assuming the corresponding scheduling matrices of $\boldsymbol{\mathcal{G}}_i$ pseudo-commute with $\mathbf{S}_i$. It will now be shown that the scheduling matrices can be constructed to satisfy the pseudo-commutativity condition in (6.8) for any matrix $\mathbf{S}_i \in \mathbb{R}^{n_y \times n_u}$.

Consider an arbitrary matrix $\mathbf{S}_i \in \mathbb{R}^{n_y \times n_u}$ and the corresponding input scheduling matrix $\mathbf{\Phi}_{u,i}(t) \in \mathbb{R}^{n_u \times n_u}$ and output scheduling matrix $\mathbf{\Phi}_{y,i}(t) \in \mathbb{R}^{n_y \times n_y}$. For the trivial case of $\mathbf{S}_i = \mathbf{0}$, the $n_u^2 + n_y^2$ entries of the input and output scheduling matrices can be independently designed to pseudo-commute with $\mathbf{S}_i = \mathbf{0}$. Now, consider an arbitrary nonzero matrix $\mathbf{S}_i \in \mathbb{R}^{n_y \times n_u}$ and its singular value decomposition (SVD) given by

$$\mathbf{S}_i = \mathbf{U}_i \begin{bmatrix} \mathbf{\Sigma}_{1,i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}_i^{\mathsf{T}}, \tag{6.9}$$

where $\mathbf{U}_i \in \mathbb{R}^{n_y \times n_y}$, $\mathbf{\Sigma}_{1,i} \in \mathbb{S}^{\varrho_i}$, $\mathbf{V}_i \in \mathbb{R}^{n_u \times n_u}$, and $\varrho_i = \operatorname{rank}(\mathbf{S}_i) \geq 1$. As shown in Lemma A.1 found in the Appendix, the scheduling matrices $\mathbf{\Phi}_{u,i}(t) \in \mathbb{R}^{n_u \times n_u}$ and $\mathbf{\Phi}_{y,i}(t) \in \mathbb{R}^{n_y \times n_y}$ satisfy the pseudo-commutativity condition in (6.8) if and only if there exists $\mathbf{Z}_{11,i}(t)$, $\mathbf{Z}_{21,i}(t)$, $\mathbf{Z}_{22,i}(t)$,

$\mathbf{W}_{21,i}(t)$, and $\mathbf{W}_{22,i}(t)$, with appropriate dimensions, such that

$$\boldsymbol{\Phi}_{u,i}(t) = \mathbf{V}_i \begin{bmatrix} \mathbf{Z}_{11,i}(t) & \mathbf{0} \\ \mathbf{Z}_{21,i}(t) & \mathbf{Z}_{22,i}(t) \end{bmatrix} \mathbf{V}_i^{\mathsf{T}}, \tag{6.10a}$$

$$\boldsymbol{\Phi}_{y,i}(t) = \mathbf{U}_i \begin{bmatrix} \boldsymbol{\Sigma}_{1,i}^{-1}\mathbf{Z}_{11,i}^{\mathsf{T}}(t)\boldsymbol{\Sigma}_{1,i} & \mathbf{0} \\ \mathbf{W}_{21,i}(t) & \mathbf{W}_{22,i}(t) \end{bmatrix} \mathbf{U}_i^{\mathsf{T}}. \tag{6.10b}$$

Therefore, for any arbitrary matrix $\mathbf{S}_i \in \mathbb{R}^{n_y \times n_u}$ with $\varrho_i = \operatorname{rank}(\mathbf{S}_i) \geq 1$, the scheduling matrices can be designed using the $n_u^2 + n_y^2 + \varrho_i^2 - \varrho_i(n_u + n_y)$ entries of the independent design variables $\mathbf{Z}_{11,i}(t)$, $\mathbf{Z}_{21,i}(t)$, $\mathbf{Z}_{22,i}(t)$, $\mathbf{W}_{21,i}(t)$, and $\mathbf{W}_{22,i}(t)$ to satisfy the pseudo-commutativity condition in (6.8).

It is also useful to consider the implications of the pseudo-commutativity condition in (6.8) for the special case of a square matrix $\mathbf{S}_i \in \mathbb{R}^{n \times n}$. If $\mathbf{S}_i$ is square and full rank, the scheduling matrices can be designed using the similarity transformation

$$\boldsymbol{\Phi}_{u,i}(t) = \mathbf{S}_i^{-1}\boldsymbol{\Phi}_{y,i}^{\mathsf{T}}(t)\mathbf{S}_i, \qquad \forall t \in \mathbb{Z}_{\geq 0}, \forall i \in \mathcal{N}, \tag{6.11}$$

where $\boldsymbol{\Phi}_{y,i}(t) \in \mathbb{R}^{n \times n}$ are free design variables. Additionally, since similarity transformations preserve eigenvalues, the input scheduling matrix can be made full rank by requiring the output scheduling matrix to be full rank. For the case of $\mathbf{S}_i = c_i\mathbf{1}$ with $c_i \in \mathbb{R}\setminus\{0\}$, the pseudo-commutativity in (6.8) simplifies to $\boldsymbol{\Phi}_{u,i}(t) = \boldsymbol{\Phi}_{y,i}^{\mathsf{T}}(t)$, which is exactly the scheduling-matrix architecture presented in [3].

6.4 QSR-Dissipativity of the Matrix-Gain-Scheduled System

This section presents the QSR-dissipativity of the matrix-gain-scheduled system in Figure 6.1 for two cases.

Case 1: Subsystems being QSR-dissipative with $\mathbf{Q}_i \in \mathbb{S}_{--}^{n_y}$.

Case 2: Subsystems being QSR-dissipative with $\mathbf{Q}_i \in \mathbb{S}_{-}^{n_y}$ and $\mathbf{S}_i = \mathbf{S} \in \mathbb{R}^{n_y \times n_u}$.

Doing so extends the cases discussed in [21] while generalizing the scheduling approach by allowing the scheduling signals to be scheduling matrices.

6.4.1 Subsystems being QSR-Dissipative with $\mathbf{Q}_i \in \mathbb{S}_{--}^{n_y}$

The following lemma establishes the QSR-dissipativity of the gain-scheduled subsystem $\bar{\boldsymbol{\mathcal{G}}}_i$ in Figure 6.2.

Lemma 6.1. Consider the subsystem $\boldsymbol{\mathcal{G}}_i$ in Figure 6.2 being QSR-dissipative with $\mathbf{Q}_i \in \mathbb{S}_{--}^{n_y}$,

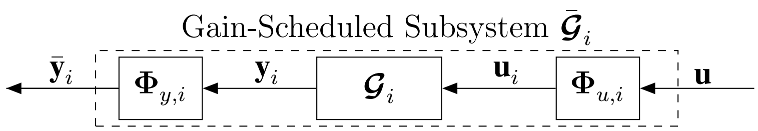


Figure 6.2: Gain-scheduling of the i^{th} subsystem, $\boldsymbol{\mathcal{G}}_i$. The input and output of $\boldsymbol{\mathcal{G}}_i$ are scheduled as per (6.3) via the matrix multiplication between the scheduling matrices $\mathbf{\Phi}_{u,i}(t)$ and $\mathbf{\Phi}_{y,i}(t)$, and their corresponding signals $\mathbf{u}(t)$ and $\mathbf{y}_i(t)$, respectively.

$\mathbf{S}_i \in \mathbb{R}^{n_y \times n_u}$, and $\mathbf{R}_i \in \mathbb{S}^{n_u}$. The discrete-time gain-scheduled subsystem $\bar{\boldsymbol{\mathcal{G}}}_i$ is QSR-dissipative, provided its corresponding scheduling matrices pseudo-commute with $\mathbf{S}_i$.

Proof. Refer to Appendix C. □

The following theorem establishes the QSR-dissipativity of the gain-scheduled system, $\bar{\boldsymbol{\mathcal{G}}}$, in Figure 6.1 where each subsystem $\boldsymbol{\mathcal{G}}_i$ is QSR-dissipative with $\mathbf{Q}_i \in \mathbb{S}^{n_y}_{--}$.

Theorem 6.1. Given that each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ is QSR-dissipative with $\mathbf{Q}_i \in \mathbb{S}^{n_y}_{--}$, $\mathbf{S}_i \in \mathbb{R}^{n_y \times n_u}$, and $\mathbf{R}_i \in \mathbb{S}^{n_u}$, the discrete-time gain-scheduled system $\bar{\boldsymbol{\mathcal{G}}}$ in Figure 6.1 is QSR-dissipative, provided the scheduling matrices $\mathbf{\Phi}_{u,i}(t)$ and $\mathbf{\Phi}_{y,i}(t)$ pseudo-commute with $\mathbf{S}_i$.

Proof. Refer to Appendix C. □

6.4.2 Subsystems being QSR-Dissipative with $\mathbf{Q}_i \in \mathbb{S}^{n_y}_{-}$ and $\mathbf{S}_i = \mathbf{S} \in \mathbb{R}^{n_y \times n_u}$

The following lemma establishes the resulting $\mathbf{Q}$ for the gain-scheduled system.

Lemma 6.2. Given each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ is QSR-dissipative with $\mathbf{Q}_i \in \mathbb{S}^{n_y}_{-}$, there exists a $\mathbf{Q} = -\varepsilon \mathbf{1}$ with $\varepsilon \in \mathbb{R}_{\geq 0}$ such that the discrete-time gain-scheduled system $\bar{\boldsymbol{\mathcal{G}}}$ in Figure 6.1 satisfies $\sum_{i \in \mathcal{N}} \langle \mathbf{y}_i, \mathbf{Q}_i \mathbf{y}_i \rangle_T \leq \langle \mathbf{y}, \mathbf{Q}\mathbf{y} \rangle_T$, provided the output scheduling matrices are active.

Proof. Refer to Appendix C. □

By defining the index sets

$$\mathcal{R}_{>0} = \{\, i \in \mathcal{N} \mid \lambda_{\max}(\mathbf{R}_i) > 0 \,\}, \tag{6.12a}$$

$$\mathcal{R}_{<0} = \{\, i \in \mathcal{N} \mid \lambda_{\max}(\mathbf{R}_i) < 0 \,\}, \tag{6.12b}$$

$$\mathcal{R}_{0} = \{\, i \in \mathcal{N} \mid \lambda_{\max}(\mathbf{R}_i) = 0 \,\}, \tag{6.12c}$$

the following lemma establishes the resulting $\mathbf{R}$ for the gain-scheduled system.

Lemma 6.3. Given each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ is QSR-dissipative with $\mathbf{R}_i \in \mathbb{S}^{n_u}$, there exists an $\mathbf{R} = \delta \mathbf{1}$ with $\delta \in \mathbb{R}$ such that the discrete-time gain-scheduled system, $\bar{\boldsymbol{\mathcal{G}}}$, in

Figure 6.1 satisfies $\sum_{i\in\mathcal{N}} \langle \mathbf{u}_i, \mathbf{R}_i\mathbf{u}_i\rangle_T \leq \langle \mathbf{u}, \mathbf{R}\mathbf{u}\rangle_T$.

Proof. Refer to Appendix C. $\square$

Since Lemma 6.3 considers the most general case of subsystems being QSR-dissipative with $\mathbf{R}_i \in \mathbb{S}^{n_u}$ and imposes no further restriction on the eigenvalues of $\mathbf{R}_i$, the sign of δ in (C.36c) cannot be determined without additional information.

Corollary 6.1. Consider the discrete-time gain-scheduled system, $\bar{\boldsymbol{\mathcal{G}}}$, in Figure 6.1 where each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ is QSR-dissipative. Given the sets $\mathcal{R}_{>0}$, $\mathcal{R}_{<0}$, and $\mathcal{R}_0$ defined in (6.12), the special cases of Lemma 6.3 are as follows:

1. $\delta \in \mathbb{R}_{\geq 0}$ provided $\mathcal{R}_{>0} \neq \varnothing$ and $\mathcal{R}_{<0} = \varnothing$.
   1.1. $\delta \in \mathbb{R}_{>0}$, provided further $\exists t \in \mathcal{T}$ and an $i \in \mathcal{R}_{>0}$ such that $\mathbf{\Phi}_{u,i}(t) \neq \mathbf{0}$.
2. $\delta \in \mathbb{R}_{\leq 0}$ provided $\mathcal{R}_{>0} = \varnothing$ and $\mathcal{R}_{<0} \neq \varnothing$.
   2.1. $\delta \in \mathbb{R}_{<0}$, provided further $\mathcal{R}_{<0} \cap \mathcal{I}_u(t) \neq \varnothing$, $\forall t \in \mathcal{T}$, meaning for $i \in \mathcal{R}_{<0}$, the nonempty subset of input scheduling matrices, $\mathbf{\Phi}_{u,i}(t)$, is strongly active.
3. $\delta = 0$ provided $\mathcal{R}_{>0} = \varnothing$ and $\mathcal{R}_{<0} = \varnothing$.

Proof. The proof for each case follows from the definition of δ in (C.36c) and the sets $\mathcal{R}_{>0}$, $\mathcal{R}_{<0}$, and $\mathcal{R}_0$ in (6.12). The strictly positive case follows from the fact that $\bar{\sigma}_u \in \mathbb{R}_{>0}$, provided there exists a $t \in \mathbb{Z}_{\geq 0}$ and an $i \in \mathcal{R}_{>0}$ such that $\mathbf{\Phi}_{u,i}(t) \neq \mathbf{0}$. The strictly negative case follows from

$$\bar{\nu}_u = \inf_{t\in\mathbb{Z}_{\geq 0}} \sum_{i\in\mathcal{R}_{<0}} \nu_{u,i}^2(t) = \inf_{t\in\mathbb{Z}_{\geq 0}} \sum_{i\in\mathcal{R}_{<0}\cap\mathcal{I}_u(t)} \nu_{u,i}^2(t) > 0,$$

provided that $\mathcal{R}_{<0} \cap \mathcal{I}_u(t) \neq \varnothing$ for all $t \in \mathbb{Z}_{\geq 0}$. $\square$

Theorem 6.2. Given each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ is QSR-dissipative with $\mathbf{Q}_i \in \mathbb{S}_{-}^{n_y}$, $\mathbf{S}_i = \mathbf{S} \in \mathbb{R}^{n_y\times n_u}$, and $\mathbf{R}_i \in \mathbb{S}^{n_u}$, the discrete-time gain-scheduled system, $\bar{\boldsymbol{\mathcal{G}}}$, in Figure 6.1 is QSR-dissipative, provided the output scheduling matrices are active, and the scheduling matrices pseudo-commute with $\mathbf{S}$.

Proof. Summing (6.5) over all N subsystems yields

$$\sum_{i\in\mathcal{N}} \tilde{V}_i \leq \sum_{i\in\mathcal{N}} \langle \mathbf{y}_i, \mathbf{Q}_i\mathbf{y}_i\rangle_T + 2\sum_{i\in\mathcal{N}} \langle \mathbf{y}_i, \mathbf{S}\mathbf{u}_i\rangle_T + \sum_{i\in\mathcal{N}} \langle \mathbf{u}_i, \mathbf{R}_i\mathbf{u}_i\rangle_T . \tag{6.13}$$

Consider the term $\sum_{i\in\mathcal{N}} \langle \mathbf{y}_i, \mathbf{S}\mathbf{u}_i\rangle_T$. Given the scheduling matrices pseudo-commute with respect to $\mathbf{S}$ such that $\mathbf{\Phi}_{y,i}^{\mathsf{T}}(t)\mathbf{S} = \mathbf{S}\mathbf{\Phi}_{u,i}(t)$, and using (6.3) and (6.4), it follows that

$$\begin{aligned}\sum_{i\in\mathcal{N}} \langle \mathbf{y}_i, \mathbf{S}\mathbf{u}_i \rangle_T = \sum_{i\in\mathcal{N}} \left\langle \mathbf{y}_i, \mathbf{S}\mathbf{\Phi}_{u,i}\mathbf{u} \right\rangle_T &= \sum_{i\in\mathcal{N}} \left\langle \mathbf{y}_i, \mathbf{\Phi}_{y,i}^{\mathsf{T}}\mathbf{S}\mathbf{u} \right\rangle_T \\ &= \sum_{i\in\mathcal{N}} \left\langle \mathbf{\Phi}_{y,i}\mathbf{y}_i, \mathbf{S}\mathbf{u} \right\rangle_T = \langle \mathbf{y}, \mathbf{S}\mathbf{u} \rangle_T . \end{aligned} \tag{6.14}$$

Consequently, using Lemmas 6.2 and 6.3 and substituting (6.14) into (6.13) yields

$$\tilde{V} = \sum_{i\in\mathcal{N}} \tilde{V}_i \leq \langle \mathbf{y}, \mathbf{Q}\mathbf{y} \rangle_T + 2 \langle \mathbf{y}, \mathbf{S}\mathbf{u} \rangle_T + \langle \mathbf{u}, \mathbf{R}\mathbf{u} \rangle_T ,$$

where $\mathbf{Q} = -\varepsilon\mathbf{1}$ and $\mathbf{R} = \delta\mathbf{1}$, with ε and δ defined as per (C.31) and (C.36), respectively. Therefore, the gain-scheduled system, $\bar{\boldsymbol{\mathcal{G}}}$, is QSR-dissipative with $\mathbf{Q} \in \mathbb{S}_{-}^{n_y}$, $\mathbf{S} \in \mathbb{R}^{n_y \times n_u}$, and $\mathbf{R} \in \mathbb{S}^{n_u}$. $\square$

6.5 Matrix-Scheduling of Special Cases of QSR-Dissipative Systems

In this section, the special cases of QSR-dissipativity, as defined in Definition 6.1, are discussed for the gain-scheduled system $\bar{\boldsymbol{\mathcal{G}}}$ in Figure 6.1. For each case, to compare the discrete-time matrix-scheduling results presented in Theorems 6.1 and 6.2 with the existing literature on scalar scheduling signals, the special case of $\mathbf{\Phi}_{u,i}(t) = s_i(t)\mathbf{1}$ and $\mathbf{\Phi}_{y,i}(t) = s_i(t)\mathbf{1}$ for all $i \in \mathcal{N}$ is considered and referred to as the *base case*.

Moreover, the matrix-scheduling architecture in [3] includes additional constant scaling parameters, $\alpha_i \in \mathbb{R}_{>0}$, for each subsystem $\boldsymbol{\mathcal{G}}_i$. These scaling parameters are not considered in this work to maintain focus on the proposed generalized matrix-gain-scheduling architecture. However, they can easily be incorporated into the results presented in Theorems 6.1 and 6.2. Therefore, when comparing the results of this work with [3], the scaling parameters are assumed to be $\alpha_i = 1$ for all $i \in \mathcal{N}$.

Finally, when discussing the results of [2–4, 17–22, 50], it is assumed that truncated signals and truncated inner products are properly defined for $t \in [0, T]$, for all $T \in \mathbb{R}_{\geq 0}$.

6.5.1 Passive Subsystems

Consider the case where each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ is passive with $\mathbf{Q}_i = \mathbf{0}$, $\mathbf{S}_i = \frac{1}{2}\mathbf{1}$, and $\mathbf{R}_i = \mathbf{0}$. Theorem 6.2 suggests $\bar{\boldsymbol{\mathcal{G}}}$ is QSR-dissipative with $\mathbf{Q} = \mathbf{0}, \mathbf{S} = \frac{1}{2}\mathbf{1}$, and $\mathbf{R} = \mathbf{0}$, and therefore is passive, provided the scheduling matrices pseudo-commute with $\mathbf{S}$. Therefore, the gain-scheduling of passive subsystems using scheduling matrices results in an overall passive gain-scheduled system. For the base case, this result is consistent with the results shown in [25, Theorem 8.2], which is again shown in [21, Theorem 2] and [22, Corollary 6.1.2].

Note, Theorem 6.2 further requires the output scheduling matrices to be active. However, this condition can be relaxed for the passive case since $\mathbf{Q}_i = \mathbf{0}$, for all $i \in \mathcal{N}$. As discussed in Section 6.3.3, for this choice of $\mathbf{S} = \frac{1}{2}\mathbf{1}$, the pseudo-commutativity condition leads to $\mathbf{\Phi}_{u,i}(t) = \mathbf{\Phi}^{\mathsf{T}}_{y,i}(t)$, recovering the matrix-scheduling architecture presented in [3].

6.5.2 Input Strictly Passive Subsystems

Consider the case where each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ is ISP with $\mathbf{Q}_i = \mathbf{0}$, $\mathbf{S}_i = \frac{1}{2}\mathbf{1}$, and $\mathbf{R}_i = -\delta_i \mathbf{1}$, where $\delta_i \in \mathbb{R}_{>0}$. Theorem 6.2 suggests $\bar{\boldsymbol{\mathcal{G}}}$ is QSR-dissipative with

$$\mathbf{Q} = \mathbf{0}, \qquad \mathbf{S} = \frac{1}{2}\mathbf{1}, \qquad \mathbf{R} = -\delta \mathbf{1},$$

where

$$\delta = \min_{i \in \mathcal{N}}(\delta_i) \inf_{t \in \mathbb{Z}_{\geq 0}} \sum_{i \in \mathcal{N}} \nu^2_{u,i}(t) \geq 0, \tag{6.15}$$

provided the scheduling matrices pseudo-commute with $\mathbf{S} = \frac{1}{2}\mathbf{1}$. Furthermore, if the scheduling matrices are designed to be strongly active, Corollary 6.1 with $\mathcal{R}_{<0} = \mathcal{N}$ suggests $\bar{\boldsymbol{\mathcal{G}}}$ is ISP with

$$\delta = \delta_{\min} \inf_{t \in \mathbb{Z}_{\geq 0}} \sum_{i \in \mathcal{I}_u(t)} \nu^2_{u,i}(t) > 0, \tag{6.16}$$

where $\delta_{\min} = \min_{i \in \mathcal{N}} \delta_i$. Note, as per the definition of $\mathcal{I}_u(t)$ in (6.6), $s_i(t) = 0$ for $i \in \mathcal{N} \setminus \mathcal{I}_u(t)$. The ISP coefficient in (6.16) is the discrete-time analogous of the matrix scheduling result for ISP subsystems presented in [3, Theorem 1]. For the base case, the ISP coefficient δ in (6.16) simplifies to

$$\delta = \delta_{\min} \inf_{t \in \mathbb{Z}_{\geq 0}} \sum_{i \in \mathcal{I}_u(t)} |s_i(t)|^2, \tag{6.17}$$

which is again, the discrete-time analogous of the results in [2, Theorem 1].

Contrary to [21], at any time, [2, Theorem 1] requires at least one scheduling signal to be nonzero. This ensures that for all time $t \in \mathbb{R}_{\geq 0}$, there exists at least one $i \in \mathcal{N}$ such that $\mathbf{u}_i(t)$ and $\bar{\mathbf{y}}_i(t)$ are nonzero, provided that $\mathbf{u}(t)$ and $\mathbf{y}_i(t)$ are nonzero. Here, the scheduling matrices are required to pseudo-commute with $\mathbf{S}$ and be strongly active to ensure for all time $t \in \mathbb{Z}_{\geq 0}$, there exists at least one $i \in \mathcal{N}$ such that $\mathbf{\Phi}_{u,i}(t) = \mathbf{\Phi}^{\mathsf{T}}_{y,i}(t)$ is full rank. The existence of a full rank scheduling matrix at each time ensures that there exists at least one $i \in \mathcal{N}$ such that $\mathbf{u}_i(t)$ and $\bar{\mathbf{y}}_i(t)$ are nonzero, provided that $\mathbf{u}(t)$ and $\mathbf{y}_i(t)$ are nonzero, which can be thought of as direct extension of the aforementioned condition in [2, Theorem 1].

6.5.3 Output Strictly Passive Subsystems

Consider the case where each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ is OSP with $\mathbf{Q}_i = -\varepsilon_i \mathbf{1}$, $\mathbf{S}_i = \frac{1}{2}\mathbf{1}$, and $\mathbf{R}_i = \mathbf{0}$, where $\varepsilon_i \in \mathbb{R}_{>0}$. Theorem 6.1 suggests $\bar{\boldsymbol{\mathcal{G}}}$ is QSR-dissipative with

$$\mathbf{Q} = -\varepsilon_{\min}\mathbf{1}, \qquad \mathbf{S} = \varepsilon_{\min} \sum_{i \in \mathcal{N}} \frac{\bar{\sigma}_{y,i}^2}{2\varepsilon_i}\mathbf{1}, \qquad \mathbf{R} = \left(N \sum_{i \in \mathcal{N}} \left(\frac{\bar{\sigma}_{y,i}^4}{4\varepsilon_i} \right) - \varepsilon_{\min} \left(\sum_{i \in \mathcal{N}} \frac{\bar{\sigma}_{y,i}^2}{2\varepsilon_i} \right)^2 \right) \mathbf{1},$$

provided the scheduling matrices pseudo-commute with $\mathbf{S} = \frac{1}{2}\mathbf{1}$. Notice $\mathbf{R} \in \mathbb{S}_+^{n_u}$ as a consequence of Lemma 2.3, since

$$N \sum_{i \in \mathcal{N}} \frac{\bar{\sigma}_{y,i}^4}{4\varepsilon_i} = N \sum_{i \in \mathcal{N}} \varepsilon_i \frac{\bar{\sigma}_{y,i}^4}{4\varepsilon_i^2} \geq \left(\sum_{i \in \mathcal{N}} \sqrt{\varepsilon_i} \frac{\bar{\sigma}_{y,i}^2}{2\varepsilon_i} \right)^2 \geq \left(\min_{i \in \mathcal{N}} \left(\sqrt{\varepsilon_i}\right) \sum_{i \in \mathcal{N}} \frac{\bar{\sigma}_{y,i}^2}{2\varepsilon_i} \right)^2 = \varepsilon_{\min} \left(\sum_{i \in \mathcal{N}} \frac{\bar{\sigma}_{y,i}^2}{2\varepsilon_i} \right)^2,$$

with $\mathbf{R} = \mathbf{0}$ if and only if $\bar{\sigma}_{y,i}^2 / \sqrt{\varepsilon_i} = k$, for some positive constant $k \in \mathbb{R}_{>0}$ and for all $i \in \mathcal{N}$. Therefore, Theorem 6.1 does not provide a tight bound on $\mathbf{R}$ for the OSP case. Similarly, for the base case, [21, Theorem 1] also leads to $\mathbf{R} \in \mathbb{S}_+^{n_u}$, with $\mathbf{R} = \mathbf{0}$ if and only if $\sup_{t \in \mathbb{R}_{\geq 0}} |s_i(t)|^2 / \varepsilon_i = k$, for some positive constant $k \in \mathbb{R}_{>0}$, for all $i \in \mathcal{N}$. However, Theorem 6.2 suggests $\bar{\boldsymbol{\mathcal{G}}}$ is QSR-dissipative with

$$\mathbf{Q} = -\varepsilon\mathbf{1}, \qquad \mathbf{S} = \frac{1}{2}\mathbf{1}, \qquad \mathbf{R} = \mathbf{0},$$

and therefore is OSP with

$$\varepsilon = \frac{\varepsilon_{\min}}{\bar{\sigma}_{\boldsymbol{\Psi}}^2} > 0, \tag{6.18}$$

provided the output scheduling matrices are active and the scheduling matrices pseudo-commute with $\mathbf{S} = \frac{1}{2}\mathbf{1}$. Additionally, for the base case, the augmented matrix, $\boldsymbol{\Psi}(t)$, defined in (C.28), can be written as

$$\boldsymbol{\Psi}(t) = \begin{bmatrix} s_1(t) & \cdots & s_N(t) \end{bmatrix} \otimes \mathbf{1} = \mathbf{s}^{\mathsf{T}}(t) \otimes \mathbf{1},$$

where $\otimes$ denotes the Kronecker product. Since the output scheduling matrices are assumed to be active, then $\mathbf{s}(t) \neq \mathbf{0}$ for all $t \in \mathbb{Z}_{\geq 0}$. Therefore, the nonzero singular values of $\boldsymbol{\Psi}(t)$ are the positive numbers $\{\, \sigma(\mathbf{s}(t))\sigma_i(\mathbf{1}) \mid i \in \mathcal{N}_y \,\}$, where $\mathcal{N}_y = \{1, 2, \ldots, n_y\}$ [51, Theorem 4.2.15].

Consequently,

$$\sigma_{\boldsymbol{\Psi}}(t) = \max_{i \in \mathcal{N}_y} \sigma(\mathbf{s}(t))\sigma_i(\mathbf{1}) = \sqrt{\sum_{i \in \mathcal{N}} |s_i(t)|^2} > 0.$$

From (C.31), it follows that

$$\bar{\sigma}^2_{\boldsymbol{\Psi}} = \sup_{t \in \mathbb{Z}_{\geq 0}} \sum_{i \in \mathcal{N}} |s_i(t)|^2 > 0. \tag{6.19}$$

Similar to the ISP case, the OSP coefficient, ε, in (6.18) is the discrete-time analogous of the results in [3, Theorem 2]. For the base case, (6.19) further expands on the relationship between the scheduling signals and the OSP coefficient

$$\varepsilon = \frac{\varepsilon_{\min}}{\sup_{t \in \mathbb{Z}_{\geq 0}} \sum_{i \in \mathcal{N}} |s_i(t)|^2} > 0. \tag{6.20}$$

Using the Cauchy–Schwarz inequality, it can be shown that if a system is OSP with $\mathbf{Q} = -\varepsilon\mathbf{1}$, it also possesses finite ℓ_2 gain such that $\varepsilon = 1/\gamma$. Here, by defining $\varepsilon_i = 1/\gamma_i$ with $\gamma_i \in \mathbb{R}_{>0}$, (6.20) can be rewritten as $\varepsilon = 1/\gamma$, where

$$\gamma = \max_{i \in \mathcal{N}} (\gamma_i) \sup_{t \in \mathbb{Z}_{\geq 0}} \sum_{i \in \mathcal{N}} |s_i(t)|^2 > 0. \tag{6.21}$$

Therefore, the gain-scheduled system $\bar{\boldsymbol{\mathcal{G}}}$ possess finite ℓ_2 gain with γ defined as per (6.21).

6.5.4 Subsystems Possessing Finite ℓ_2 Gain

Consider the case where each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ has finite ℓ_2 gain with $\mathbf{Q}_i = -\mathbf{1}$, $\mathbf{S}_i = \mathbf{0}$, and $\mathbf{R}_i = \gamma_i^2\mathbf{1}$, where $\gamma_i \in \mathbb{R}_{>0}$. Theorem 6.2 suggests $\bar{\boldsymbol{\mathcal{G}}}$ is QSR-dissipative with

$$\mathbf{Q} = -\frac{1}{\bar{\sigma}^2_{\boldsymbol{\Psi}}}\mathbf{1}, \qquad \mathbf{S} = \mathbf{0}, \qquad \mathbf{R} = \bar{\sigma}_u \max_{i \in \mathcal{N}} \gamma_i^2 \mathbf{1}, \tag{6.22}$$

provided the output scheduling matrices are active. Since $\mathbf{S} = \mathbf{0}$, the pseudo-commutativity condition is trivially satisfied, effectively decoupling the input and output scheduling matrices. To get a more familiar form, (6.22) can be scaled by $\bar{\sigma}^2_{\boldsymbol{\Psi}} > 0$ to obtain

$$\mathbf{Q} = -\mathbf{1}, \qquad \mathbf{S} = \mathbf{0}, \qquad \mathbf{R} = \bar{\sigma}^2_{\boldsymbol{\Psi}} \bar{\sigma}_u \max_{i \in \mathcal{N}} \gamma_i^2 \mathbf{1}. \tag{6.23}$$

Considering the base case and Corollary 6.1 with $\mathcal{R}_{>0} = \mathcal{N}$, substituting (C.36a) and (6.19) into (6.23) leads to

$$\mathbf{Q} = -\mathbf{1}, \qquad \mathbf{S} = \mathbf{0}, \qquad \mathbf{R} = \gamma^2\mathbf{1},$$

where $\bar{\mathcal{G}}$ possess finite ℓ_2 gain with the exact same γ derived in (6.21), where the subsystems were assumed to be OSP. Moreover, the gain in (6.21) is similar to that of [17, Theorem 5.2]. It is unclear which gain is more conservative since here, the maximum γ_i over all subsystems is used, whereas in [17, Theorem 5.2], the γ_i terms remain in the summation. On the other hand, (6.21) uses the supremum of sums of scheduling signals, $s_i(t)$, whereas [17, Theorem 5.2] uses sum of suprema of scheduling signals. Alternatively, Theorem 6.1 suggests $\bar{\mathcal{G}}$ is QSR-dissipative with $\mathbf{Q} = -\mathbf{1}$, $\mathbf{S} = \mathbf{0}$, and

$$\mathbf{R} = \left(N \sum_{i \in \mathcal{N}} \gamma_i^2 \bar{\sigma}_{y,i}^2 \bar{\sigma}_{u,i}^2 \right) \mathbf{1}. \tag{6.24}$$

Again, for the base case, (6.24) simplifies to

$$\mathbf{R} = \left(N \sum_{i \in \mathcal{N}} \gamma_i^2 \|s_i\|_\infty^4 \right) \mathbf{1}, \tag{6.25}$$

where $\|s_i\|_\infty = \sup_{t \in \mathbb{Z}_{\geq 0}} |s_i(t)|$. Furthermore, extending [21, Theorem 1] to the finite ℓ_2 gain case recovers (6.25). Using the inequality in Lemma 2.3, it can be shown that

$$N \sum_{i \in \mathcal{N}} \gamma_i^2 \|s_i\|_\infty^4 \geq \left(\sum_{i \in \mathcal{N}} \gamma_i \|s_i\|_\infty^2 \right)^2,$$

meaning, [17, Theorem 5.2] is less conservative than Theorem 6.1 and [21, Theorem 1] for the finite ℓ_2 gain case.

6.5.5 Very Strictly Passive Subsystems

Consider the case where each subsystem $\mathcal{G}_i$ for $i \in \mathcal{N}$ is VSP with $\mathbf{Q}_i = -\varepsilon_i \mathbf{1}$, $\mathbf{S}_i = \frac{1}{2}\mathbf{1}$, and $\mathbf{R}_i = -\delta_i \mathbf{1}$, where $\varepsilon_i, \delta_i \in \mathbb{R}_{>0}$. Theorem 6.2 suggests $\bar{\mathcal{G}}$ is QSR-dissipative with

$$\mathbf{Q} = -\varepsilon \mathbf{1}, \qquad \mathbf{S} = \frac{1}{2}\mathbf{1}, \qquad \mathbf{R} = -\delta \mathbf{1},$$

with $\varepsilon \in \mathbb{R}_{>0}$ and $\delta \in \mathbb{R}_{\geq 0}$ given by (6.18) and (6.15), respectively, provided the output scheduling matrices are active and the scheduling matrices pseudo-commute with $\mathbf{S} = \frac{1}{2}\mathbf{1}$. Furthermore, if the scheduling matrices are strongly active, Corollary 6.1 with $\mathcal{R}_{<0} = \mathcal{N}$ suggests $\bar{\mathcal{G}}$ is VSP with $\delta \in \mathbb{R}_{>0}$ given in (6.16). In [17], the authors combine the ISP and finite $\mathcal{L}_2$ gain results to show that the gain-scheduled system is VSP, while in [3], the ISP and OSP results are combined to achieve the same. Similar to the OSP case, Theorem 6.1 and [21, Theorem 1] can still be used to show the overall gain-scheduled system $\bar{\boldsymbol{\mathcal{G}}}$ is QSR-dissipative. However, they cannot ensure $\bar{\boldsymbol{\mathcal{G}}}$ is VSP as they lead to $\delta \in \mathbb{R}$ without the further guarantee of $\delta > 0$.

6.5.6 Conic Subsystems

Consider the case where each subsystem $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{N}$ is conic with $\mathbf{Q}_i = -\mathbf{1}$, $\mathbf{S}_i = c_i\mathbf{1}$, and $\mathbf{R}_i = (r_i^2 - c_i^2)\mathbf{1}$, where $c_i \in \mathbb{R}$ and $r_i \in \mathbb{R}_{>0}$, provided the scheduling matrices pseudo-commute with $\mathbf{S}_i$. Theorem 6.1 suggests $\bar{\boldsymbol{\mathcal{G}}}$ is QSR-dissipative with

$$\mathbf{Q} = -\mathbf{1}, \qquad \mathbf{S} = c\mathbf{1}, \qquad \mathbf{R} = \left(r^2 - c^2\right)\mathbf{1}, \tag{6.26}$$

where

$$c = \sum_{i \in \mathcal{N}} \left(c_i \bar{\sigma}_{y,i}^2\right), \tag{6.27a}$$

$$r = \sqrt{N \sum_{i \in \mathcal{N}} \left(\delta_i + \bar{\sigma}_{y,i}^4 c_i^2\right)} = \sqrt{N \sum_{i \in \mathcal{N}} \bar{r}_i^2}, \tag{6.27b}$$

are the conic center and radius, respectively, and $\bar{r}_i^2 = \delta_i + \bar{\sigma}_{y,i}^4 c_i^2$. Following the definition of δ_i in (C.9),

$$\bar{r}_i = \begin{cases} \bar{\sigma}_{y,i}\sqrt{\bar{\sigma}_{u,i}^2(r_i^2 - c_i^2) + \bar{\sigma}_{y,i}^2 c_i^2}, & \text{if } r_i > |c_i|, \\ \bar{\sigma}_{y,i}\sqrt{\bar{\nu}_{u,i}^2(r_i^2 - c_i^2) + \bar{\sigma}_{y,i}^2 c_i^2}, & \text{if } r_i \leq |c_i|. \end{cases} \tag{6.28}$$

Additionally, the scheduling matrices pseudo-commute with $\mathbf{S}_i = c_i\mathbf{1}$, meaning $\mathbf{\Phi}_{y,i}^{\mathsf{T}}(t)\mathbf{S}_i = \mathbf{S}_i\mathbf{\Phi}_{u,i}(t)$, for all $t \in \mathbb{Z}_{\geq 0}$ and $i \in \mathcal{N}$. It follows that

$$\mathbf{\Phi}_{y,i}^{\mathsf{T}}(t) = \mathbf{S}_i\mathbf{\Phi}_{u,i}(t)\mathbf{S}_i^{-1} = \mathbf{\Phi}_{u,i}(t),$$

for all $t \in \mathbb{Z}_{\geq 0}$ and $i \in \mathcal{N}$, such that $c_i \neq 0$. Therefore, the scheduling matrices corresponding to the i^{th} subsystem must share the same eigenvalues and singular values, provided $c_i \neq 0$. Consequently, (6.28) can be simplified as

$$\bar{r}_i = \begin{cases} \bar{\sigma}_{y,i}\bar{\sigma}_{u,i} r_i, & \text{if } r_i > |c_i| = 0, \\ \bar{\sigma}_i^2 r_i, & \text{if } r_i > |c_i| \neq 0, \\ \bar{\sigma}_i\sqrt{\bar{\nu}_i^2 r_i^2 + (\bar{\sigma}_i^2 - \bar{\nu}_i^2)c_i^2}, & \text{if } r_i \leq |c_i|, \end{cases} \tag{6.29}$$

where $\bar{\sigma}_i = \bar{\sigma}_{u,i} = \bar{\sigma}_{y,i}$ and $\bar{\nu}_i = \bar{\nu}_{u,i} = \bar{\nu}_{y,i}$ for all $i \in \mathcal{N}$ such that $c_i \neq 0$. For the base case, the conic center defined per (6.27a) simplifies to $c = \sum_{i\in\mathcal{N}} c_i \|s_i\|_\infty^2$, matching the results in [21, Theorem 1] and [22, Theorem 6.2.1]. Moreover, for the base case, $\bar{r}_i$ in (6.29) simplifies to the results in [21, Theorem 1] as

$$\bar{r}_i = \begin{cases} \sigma_i^2 r_i, & \text{if } r_i > |c_i|, \\ \sigma_i \sqrt{\nu_i^2 r_i^2 + (\sigma_i^2 - \nu_i^2) c_i^2}, & \text{if } r_i \leq |c_i|, \end{cases}$$

with $\sigma_i = \|s_i\|_\infty < \infty$ and $\nu_i = \inf_{t\in\mathbb{Z}_{\geq 0}} |s_i(t)| \geq 0$, for all $i \in \mathcal{N}$. As discussed in [21], when $r_i > |c_i|$, [22, Theorem 6.2.1] provides a less conservative bound on the conic radius of the gain-scheduled system compared to Theorem 6.1. This is again as a direct result of the special case of Cauchy–Schwarz inequality used in the proof of Theorem 6.1. Alternatively, since $\nu_i \leq \sigma_i$ for all $i \in \mathcal{N}$, Theorem 6.1 provides a tighter bound on the conic radius of the gain-scheduled system when $r_i \leq |c_i|$ compared to [22, Theorem 6.2.1].

6.6 Applications of Gain-Scheduling of First-Order Optimization Algorithms

In Chapter 4 and Chapter 5, it was shown that by performing a loop transformation, the first-order optimization algorithms can be interpreted as passive controllers in negative feedback with a modified gradient. In Section 6.5.1, it was shown that the gain-scheduling of passive subsystems results in an overall passive gain-scheduled system. By combining these two results, this section proposes a control interpretation of first-order optimization algorithms with varying hyperparameters as gain-scheduled systems. Moreover, a new variation of the GD algorithm is proposed, where both the input and output of the gradient are scaled. Finally, the gain-scheduling of passive first-order optimization algorithms is discussed, where the scheduling signals can switch between different first-order optimization algorithms, such as GD, HB, NAG, and TM methods.

6.6.1 Gradient Descent with Varying Step Size as a Gain-Scheduled System

This subsection provides a gain-scheduling interpretation of the GD algorithm with varying step size using the gain-scheduling architecture presented in Section 6.3.

Given $f \in \mathcal{S}_{0,L}$ with $0 < L$, consider the linear parameter-varying (LPV) GD controller with state-space realization

$$\boldsymbol{\xi}^{k+1} = \boldsymbol{\xi}^k + \alpha^k \mathbf{u}^k, \qquad \mathbf{y}^k = \boldsymbol{\xi}^k, \tag{6.30}$$

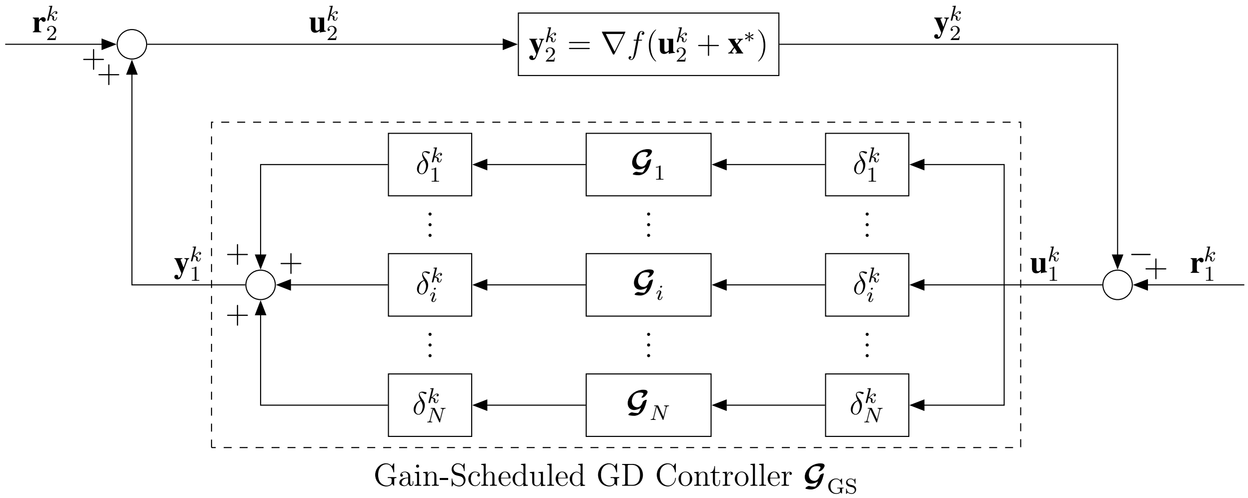


Figure 6.3: LPV GD controller as a gain-scheduled system composed of N parallel GD controllers with fixed step sizes α_i. The input and output of each controller is scaled by the Kronecker delta scheduling function δ_i^k, which is defined as $\delta_i^k = 1$ if $i = k$ and $\delta_i^k = 0$ otherwise.

where $\alpha^k \in \left(0, \frac{2}{L}\right)$ for $k \in \mathcal{T} = \{1, 2, \ldots, N\}$ and $0 < N$. Construct N GD controllers, $\boldsymbol{\mathcal{G}}_i$, with fixed step sizes α_i, such that $\alpha^k = \sum_{i \in \mathcal{T}} \delta_i^k \alpha_i$, for all $k \in \mathcal{T}$, where

$$\delta_i^k = \begin{cases} 1, & i = k, \\ 0, & i \neq k. \end{cases} \tag{6.31}$$

The LPV GD controller in (6.30) can then be represented using the gain-scheduled GD controller $\boldsymbol{\mathcal{G}}_{\text{GD}}$ in Figure 6.3, for $k \in \mathcal{T}$.

Now, consider gain-scheduling N modified GD controllers $\bar{\boldsymbol{\mathcal{G}}}_i$, where the loop transformation is applied to each controller $\boldsymbol{\mathcal{G}}_i$ in Figure 6.3 as shown in Figure 6.4. To retain the same closed-loop behavior of Figure 6.3, the feedback term $D^k \mathbf{1}$ is introduced, where

$$D^k = \sum_{i \in \mathcal{T}} \left(\delta_i^k\right)^2 D_i = \sum_{i \in \mathcal{T}} \delta_i^k D_i \tag{6.32}$$

Moreover, $\bar{\mathbf{r}}_2^k = \mathbf{r}_2^k - \sum_{i \in \mathcal{T}} \delta_i^k \mathbf{r}_1^k = \mathbf{r}_2^k - \mathbf{r}_1^k$, since $\sum_{i \in \mathcal{T}} \delta_i^k = 1$ for all $k \in \mathcal{T}$. Following Lemma 4.1, each modified GD controller $\bar{\boldsymbol{\mathcal{G}}}_i$ is passive if and only if

$$0 < \frac{\alpha_i}{2} \leq D_i. \tag{6.33}$$

As such, following Section 6.5.1, Theorem 6.2 ensures that the overall gain-scheduled modified GD controller $\bar{\boldsymbol{\mathcal{G}}}_{\text{GS}}$ is passive, provided (6.33) holds for all $i \in \mathcal{T}$. Moreover, for $f \in \mathcal{S}_{0,L}$ with

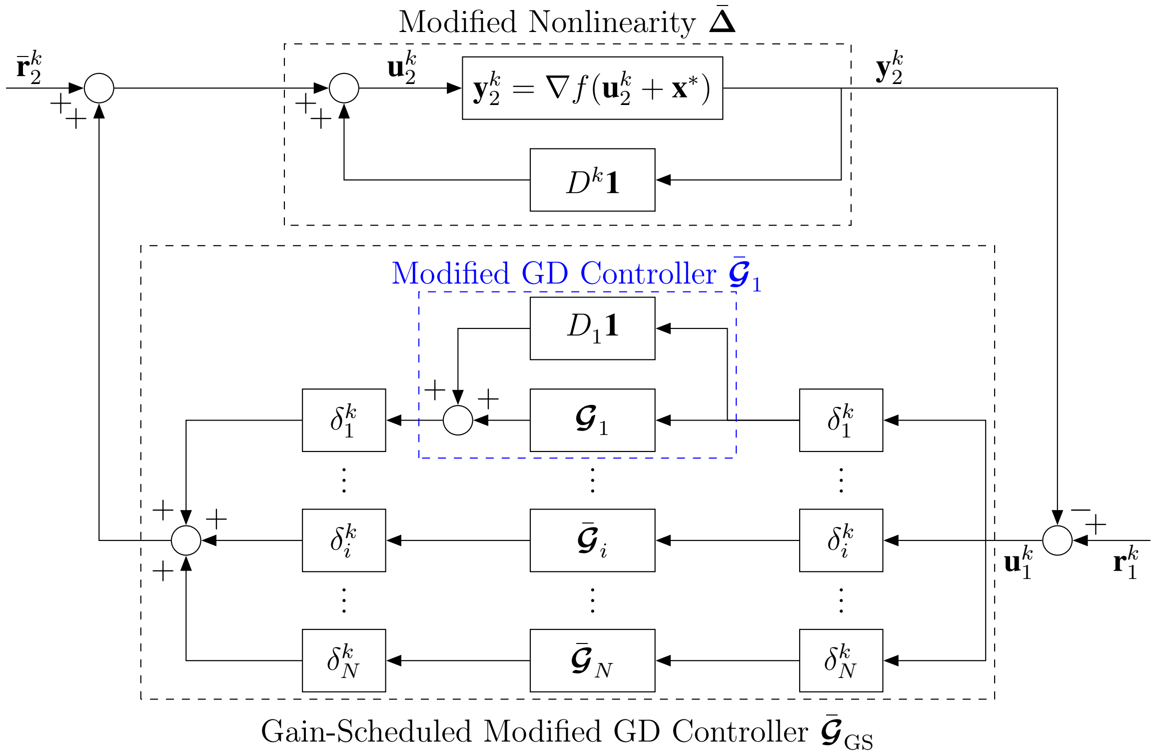


Figure 6.4: Loop transformation of the gain-scheduled GD controller in Figure 6.3. The loop transformation introduces the feedthrough terms $D_i\mathbf{1}$, resulting in the modified GD controllers $\bar{\boldsymbol{\mathcal{G}}}_i$. Moreover, the feedback term $D^k\mathbf{1}$ is introduced, resulting in the modified nonlinearity $\bar{\boldsymbol{\Delta}}$.

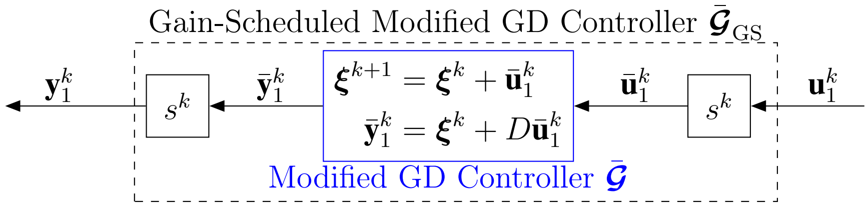
$0 < L$, Lemma 4.5 ensures that the modified nonlinearity $\bar{\boldsymbol{\Delta}}$ is OSP, provided

$$\max_{k\in\mathcal{T}} D^k = \max_{i\in\mathcal{T}} D_i < \frac{1}{L}. \tag{6.34}$$

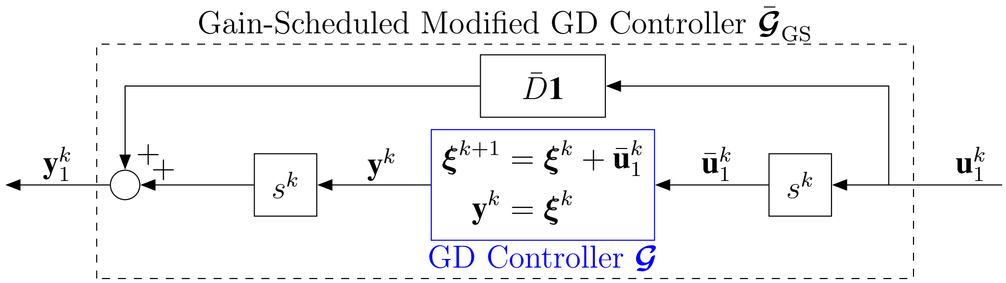
Combining (6.33) and (6.34), for $D_i = \alpha_i/2$, it follows that if

$$\max_{i\in\mathcal{T}} \alpha_i < \frac{2}{L}, \tag{6.35}$$

then $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GS}}$ is passive and $\bar{\boldsymbol{\Delta}}$ is OSP. Consequently, from Theorem 3.1-(ii), it follows that if $\mathbf{r}_1 = \mathbf{0}$ and $\bar{\mathbf{r}}_2 = \mathbf{r}_2 \in \ell_2$, then $\mathbf{y}_2 \in \ell_2$, and therefore, $\nabla f(\mathbf{x}^\mathbf{k}) \to \mathbf{0}$. As such, the proposed passivity-based analysis in thesis can also be used to guarantee the convergence of GD algorithm with varying step sizes, as long as the step sizes are bounded by $\max_{i\in\mathcal{T}} \alpha_i < \frac{2}{L}$.



(a) Gain-scheduling a modified GD controller $\bar{\boldsymbol{\mathcal{G}}}$ with step size $\alpha = 1$. The scheduling function s^k is used to scale the input and output of $\bar{\boldsymbol{\mathcal{G}}}$ such that $\bar{\mathbf{u}}_1^k = s^k\mathbf{u}_1^k$ and $\mathbf{y}_1^k = s^k\bar{\mathbf{y}}_1^k$.



(b) Gain-scheduling a GD controller $\boldsymbol{\mathcal{G}}$ with step size $\alpha = 1$. The scheduling function s^k is used to scale the input and output of $\boldsymbol{\mathcal{G}}$. A feedthrough term $\bar{D}\mathbf{1} = (s^k)^2 D\mathbf{1}$ is added such that $\mathbf{y}_1^k = \bar{D}\mathbf{u}_1^k + s^k\mathbf{y}^k$.

Figure 6.5: Two gain-scheduling architectures resulting in the same gain-scheduled modified GD controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GS}}$.

6.6.2 A New Variation of the Gradient Descent Algorithm

In Section 6.6.1, to present a gain-scheduling interpretation of the LPV GD controller, the Kronecker delta scheduling function δ_i^k was used to scale the input of the N parallel GD controllers. This subsection proposes a new variation of the GD algorithm, where both the input and output of a single GD controller are scaled by an arbitrary scheduling function s^k.

Consider the modified GD controller $\bar{\boldsymbol{\mathcal{G}}}$ in Figure 6.5a, with a fixed step size $\alpha = 1$. As per Lemma 4.1, $\bar{\boldsymbol{\mathcal{G}}}$ is passive if and only if $D \geq 1/2$. This controller is gain-scheduled as per Figure 6.2 using some general scheduling function $s^k \in \mathbb{R} \setminus 0$, for $k \in \mathcal{T} = \{0, 1, \ldots, T-1\}$, with $T \in \mathbb{Z}_{>0}$. Again, following Section 6.5.1, the gain-scheduling of a passive discrete-time system as per Figure 6.5a results in a gain-scheduled passive system. As shown in Figure 6.5b, $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GS}}$ with minimal realization $\left(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{D}}\right) = (\mathbf{1}, s^k\mathbf{1}, s^k\mathbf{1}, (s^k)^2 D\mathbf{1})$, can also be constructed by adding a feedthrough term $\bar{D}\mathbf{1} = \left(s^k\right)^2 D\mathbf{1}$ to the gain-scheduled GD controller. It follows that $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GS}}$ is passive if and only if $\bar{D} \geq \left(s^k\right)^2/2$. Moreover, the negative feedback interconnection of $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GS}}$ and $\bar{\boldsymbol{\Delta}}$, as per Figure 4.4 can then be used to represent a variation on the GD algorithm with the update rule

$$\mathbf{x}^{k+1} = \mathbf{x}^k - s^k \nabla f(s^k\mathbf{x}^k), \tag{6.36}$$

where the scheduling function $s^k \in \mathbb{R} \setminus 0$ can be used to represent time varying step sizes. Note, unlike the standard GD algorithm, where only the output of the gradient is scaled, in

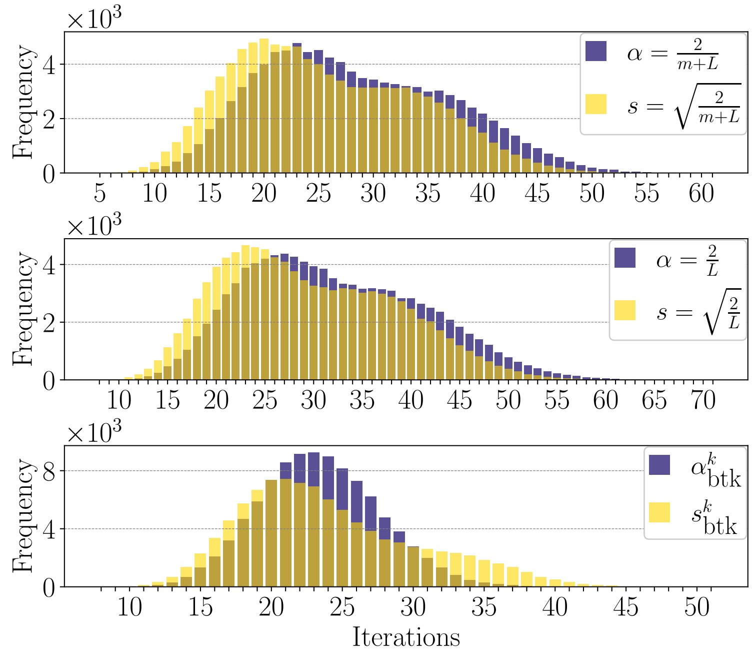


Figure 6.6: Comparison of number of iterations needed to achieve $|\nabla f(x^k)| < 10^{-12}$ for the GD method and the gain-scheduled GD method with different step sizes and scheduling functions using 10^5 uniformly sampled initial conditions for $x^0 \in [-10^5, 10^5]$.

(6.36), both the input and output of the gradient are scaled by s^k.

Remark 6.1. For a constant $s^k = s$, using the change of variable $\bar{\mathbf{x}}^k = s\mathbf{x}^k$, the update rule in (6.36) can be written as $\bar{\mathbf{x}}^{k+1} = \bar{\mathbf{x}}^k - s^2\nabla f(\bar{\mathbf{x}}^k)$, recovering the original GD algorithm with step size $\alpha = s^2$.

For $f \in \mathcal{S}_{m,L}$, $\bar{\mathbf{\Delta}}$ is VSP as per Lemma 4.3, or ISP as per Lemma 4.4, provided $\bar{D} < 1/L$ or $\bar{D} = 1/L$, respectively. Consequently, using the lower bound $\bar{D} \geq \left(s^k\right)^2/2$, a similar analysis as Theorem 4.2 shows that for $m < L$, provided $\max_{k\in\mathcal{T}}|s^k| \in (0, \sqrt{2/L}]$, then $\mathbf{y}_1^k \in \ell_2$, and therefore, $\mathbf{x}^k - D\nabla f(\mathbf{x}^k) \to \mathbf{x}^*$. As discussed in Section 4.4.1, this does not guarantee that $\nabla f(\mathbf{x}^k) \to \mathbf{0}$, but rather the auxiliary quantity $\mathbf{x}^k - D\nabla f(\mathbf{x}^k)$ converges to the unique global minimizer $\mathbf{x}^*$ of f. However, for $\max_{k\in\mathcal{T}}|s^k| \in (0, \sqrt{2/L})$, Theorem 4.2-$(i)$ guarantees that $\nabla f(\mathbf{x}^k) \to \mathbf{0}$ as $k \to \infty$.

To compare the performance of the proposed gain-scheduled GD method in (6.36) with the standard GD method, consider the unconstrained optimization problem

$$\min_{x\ in\mathbb{R}} f(x) = \frac{L-m}{4}\left(\frac{L+m}{L-m}x^2 + 2\sin(x) - 2x\cos(x)\right), \tag{6.37}$$

for $m = 1$ and $L = 100$. As discussed in Example 4.1, the function f is non-convex but has a sector-bounded gradient and a unique global minimizer at $x^* = 0$.

Table 6.1: Corresponding Mean, Median, and Mode of Figure 6.6

Parameter	Mean	Median	Mode
$\alpha = 2/(m+L)$	28.36	27	23
$s = \sqrt{2/(m+L)}$	**25.68**	**25**	**20**
$\alpha = 2/L$	32.12	31	27
$s = \sqrt{2/L}$	**29.50**	**28**	**23**
α^k_{btk}	**23.38**	23	23
s^k_{btk}	24.08	23	**21**

To solve the unconstrained optimization problem in (6.37) involving the function f, the GD method in (4.10) and the proposed gain-scheduled GD method in (6.36) are implemented for various choices of step sizes and scheduling functions. Three choices of steps sizes are considered for the GD method. The first choice is a fixed step size of $\alpha = 2/L$, being the largest possible step size to guarantee input-output stability for $m < L$, as per Theorem 4.2-(*ii*). The second choice is the optimal fixed step size of $\alpha = 2/(m+L)$ [12]. Finally, a varying step size, α^k_{btk}, is considered by implementing a line search algorithm based on Armijo's condition [52]. Similarly, for the gain-scheduled GD method in (6.36), two choices of constant scheduling functions, $s = \sqrt{2/L}$ and $s = \sqrt{2/(m+L)}$, are considered. Additionally, a varying scheduling function, s^k_{btk}, is considered by implementing a line search algorithm based on Armijo's condition with the added condition that $\max_{k\in\mathcal{T}}|s^k| \in (0, \sqrt{2/L}]$. A Monte Carlo simulation is conducted with 10^5 uniformly sampled initial conditions $x^0 \in [-10^5, 10^5]$ to compare the number of iterations needed to achieve $|\nabla f(x^k)| < 10^{-12}$ for the GD method and the gain-scheduled GD method.

As shown in Figure 6.6 and Table 6.1, for the constant step size and constant scheduling function cases, the distribution of the gain-scheduled GD method has smaller mean, median, and mode values, compared to the standard GD method. For the varying step size and varying scheduling function case, the standard GD method has a lower mean, but a higher mode, compared to the gain-scheduled GD method. The lower mode indicates better performance in terms of fewer iterations. Finally, for both methods, Table 6.1 also confirms that time-varying step sizes and scheduling functions can lead to improved performance.

6.6.3 Gain-Scheduling Passive First-Order Optimization Algorithms

In Section 6.6.1 and Section 6.6.2, the gain-scheduling of the GD algorithm was discussed. This subsection considers the gain-scheduling a mix of first-order optimization algorithms, such as GD, HB, NAG, and TM methods.

Consider the gain-scheduling architecture in Figure 6.7, where N passive first-order

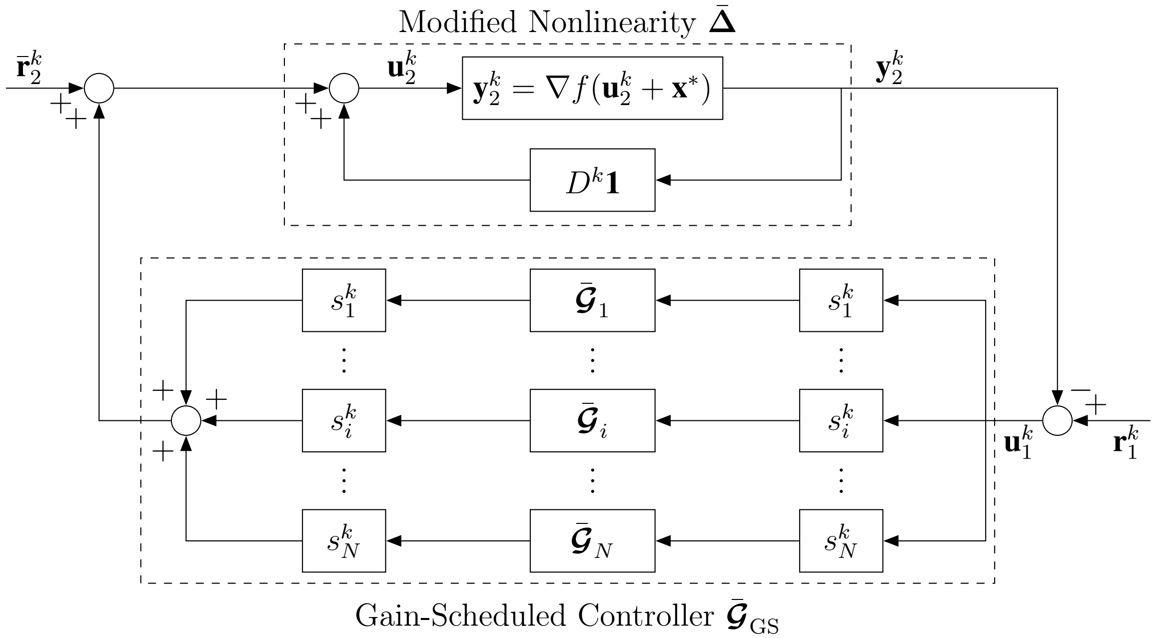


Figure 6.7: Gain-scheduling of passive first-order optimization algorithms. Each controller $\boldsymbol{\mathcal{G}}_i$, for $i \in \mathcal{N} = \{1, 2, \ldots, N\}$, can represent a different modified first-order optimization algorithm containing a feedthrough term $D_i\mathbf{1}$. The scheduling functions s_i^k are used to scale the input and output of each controller. The feedback term $D^k\mathbf{1} = \sum_{i\in\mathcal{N}}\left(s_i^k\right)^2 D_i\mathbf{1}$ is introduced to compensate for the added feedthrough terms, resulting in the modified nonlinearity $\bar{\boldsymbol{\Delta}}$.

optimization algorithms, $\bar{\boldsymbol{\mathcal{G}}}_i$, are gain-scheduled using scheduling functions s_i^k for $k \in \mathcal{T} = \{0, 1, \ldots, T-1\}$. Each controller $\boldsymbol{\mathcal{G}}_i$ can represent a different modified first-order optimization algorithm with a minimal realization $(\mathbf{A}_i, \mathbf{B}_i, \mathbf{C}_i, D_i\mathbf{1})$, where D_i is the added feedthrough term to render the controller passive. Similar to Section 6.6.1, the feedback term $D^k\mathbf{1} = \sum_{i\in\mathcal{N}}\left(s_i^k\right)^2 D_i\mathbf{1}$ is introduced to compensate for the added feedthrough terms, and $\bar{\mathbf{r}}_2^k = \mathbf{r}_2^k - \sum_{i\in\mathcal{N}}\left(s_i^k\right)^2\mathbf{r}_1^k$. For $f \in \mathcal{S}_{0,L}$ with $0 < L$, Lemma 4.5 ensures that the modified nonlinearity $\bar{\boldsymbol{\Delta}}$ is OSP, provided

$$\sup_{k\in\mathcal{T}} D^k = \sup_{k\in\mathcal{T}} \sum_{i\in\mathcal{N}} \left(s_i^k\right)^2 D_i < \frac{1}{L}. \tag{6.38}$$

As such, provided the gain-scheduled controller $\bar{\boldsymbol{\mathcal{G}}}_{\mathrm{GS}}$ is passive, $\bar{\boldsymbol{\Delta}}$ is OSP, $\mathbf{r}_1 = \mathbf{0}$, and $\bar{\mathbf{r}}_2 = \mathbf{r}_2 \in \ell_2$, then Theorem 3.1-(ii) ensures that $\nabla f(\mathbf{x}^k) \to \mathbf{0}$.

In practice, given an objective function $f \in \mathcal{S}_{0,L}$ with $0 < L$, the following steps can be used to gain-schedule first-order optimization algorithms with global convergence guarantees:

1. **Construct passive first-order optimization controllers:** Using the lower bounds provided in Table 5.2, construct passive modified first-order optimization controllers $\bar{\boldsymbol{\mathcal{G}}}_i$ for each $i \in \mathcal{N}$.

2. **Select scheduling functions:** Design the scheduling functions s_i^k for each $i \in \mathcal{N}$ such that (6.38) holds.
3. **Gain-schedule the controllers:** Gain-schedule the modified first-order optimization controllers $\bar{\boldsymbol{\mathcal{G}}}_i$ using the scheduling functions s_i^k as per Figure 6.7. It follows that the resulting gain-scheduled controller $\bar{\boldsymbol{\mathcal{G}}}_{\text{GS}}$ is passive and $\bar{\boldsymbol{\Delta}}$ is OSP. Theorem 3.1-(ii) then guarantees that $\nabla f(\mathbf{x}^k) \to \mathbf{0}$ as $k \to \infty$, for all $\mathbf{x}^0 \in \mathbb{R}^n$.

6.7 Closing Remarks

This chapter presented the discrete-time version of the matrix-scheduling of QSR-dissipative systems considered in [4]. The proposed gain-scheduling architecture was shown to result in an overall discrete-time QSR-dissipative gain-scheduled system, provided the scheduling matrices are bounded and possess certain pseudo-commutative properties. A detailed discussion was provided regarding the design and construction of scheduling matrices that satisfy the introduced pseudo-commutative property. Furthermore, the gain-scheduling of a broader class of QSR-dissipative systems was considered than that found in prior work. For the various passive cases of QSR-dissipativity, the conditions on the scheduling matrices reduced to the same conditions on the scheduling signals reported in [2, 17], while for the conic case, the results of [21] were recovered when the scheduling matrices were chosen to be a scalar times the identity matrix.

Additionally, leveraging the passivity-based analysis of first-order optimization algorithms presented in Part 1, this chapter presented a control interpretation of first-order optimization algorithms with varying hyperparameters as gain-scheduled systems. Moreover, a new variation of the GD algorithm was proposed, where both the input and output of the gradient are scaled. Finally, the gain-scheduling of passive first-order optimization algorithms was discussed, where provided the scheduling functions satisfy a certain upper bound, the global convergence of the gain-scheduled algorithm is guaranteed for $f \in \mathcal{S}_{0,L}$ with $0 < L$.

Part 3

Conclusion

Chapter 7

Closing Remarks and Future Work

This thesis presents a control-theoretic framework for analyzing and designing first-order optimization algorithms using a passivity-based gain-scheduling approach. Part 1 develops a passivity-based analysis that guarantees the input-output stability and global convergence of first-order optimization algorithms with fixed hyperparameters for a class of functions with sector-bounded gradients and a unique global minimizer. Part 2 generalizes existing gain-scheduling architectures by introducing scheduling matrices and formulating their discrete-time counterparts. By combining the results from Part 1 and Part 2, a control interpretation of first-order optimization algorithms with varying hyperparameters as gain-scheduled systems is obtained. Despite these contributions, several avenues for further research remain.

A key shortcoming of the passivity-based analysis in Part 1 is that it does not provide convergence rate guarantees. Instead, Chapter 4–5 give sufficient conditions, in the form of upper bounds on the hyperparameters, that ensure the input-output stability and global convergence of the GD, HB, NAG, and TM methods. Future work could leverage the analytic expressions for the eigenvalues of the storage function matrices in Chapter 4–5 to derive explicit convergence rate bounds. Moreover, since QSR-dissipativity is a generalization of passivity, another promising direction is to investigate stability and convergence of first-order optimization algorithms within the QSR-dissipativity framework.

In Part 2, it is shown that the matrix-scheduling of passive subsystems results in an overall passive system. Despite this general result, the applications of gain-scheduling first-order optimization algorithms are limited to the case of scalar scheduling functions. Future work could explore the use of matrix-scheduling functions to gain-schedule first-order optimization algorithms with multiple hyperparameters. Moreover, in the case of gain-scheduling optimization algorithms, such as GD, HB, NAG, and TM, the optimal choice of scheduling functions is perhaps even more ambiguous than the choice of each algorithm's respective hyperparameters.

Appendices

Appendix A

Pseudo-Commuting Matrices

Lemma A.1 ([4, Lemma 4])**.** Consider a real nonzero matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$ such that $\mathbf{S} \neq \mathbf{0}$. The arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m\times m}$ and $\mathbf{B} \in \mathbb{R}^{n\times n}$ satisfy $\mathbf{AS} = \mathbf{SB}$ if and only if

$$\mathbf{A} = \mathbf{U}\begin{bmatrix} \mathbf{\Sigma}_1\bar{\mathbf{B}}_{11}\mathbf{\Sigma}_1^{-1} & \bar{\mathbf{A}}_{12} \\ \mathbf{0} & \bar{\mathbf{A}}_{22} \end{bmatrix}\mathbf{U}^\mathsf{T}, \qquad \mathbf{B} = \mathbf{V}\begin{bmatrix} \bar{\mathbf{B}}_{11} & \mathbf{0} \\ \bar{\mathbf{B}}_{21} & \bar{\mathbf{B}}_{22} \end{bmatrix}\mathbf{V}^\mathsf{T},$$

where $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\mathbf{\Sigma}_1 \in \mathbb{S}^{\varrho}$, and $\mathbf{V} \in \mathbb{R}^{n\times n}$ are given by the SVD of $\mathbf{S}$ such that

$$\mathbf{S} = \mathbf{U}\begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^\mathsf{T},$$

with $\varrho = \operatorname{rank}(\mathbf{S}) \geq 1$, and $\bar{\mathbf{A}}_{12}, \bar{\mathbf{A}}_{22}, \bar{\mathbf{B}}_{11}, \bar{\mathbf{B}}_{21}$, and $\bar{\mathbf{B}}_{22}$ are arbitrary real matrices of appropriate dimensions.

Proof. Assume $\mathbf{A} \in \mathbb{R}^{m\times m}$ and $\mathbf{B} \in \mathbb{R}^{n\times n}$ satisfy $\mathbf{AS} = \mathbf{SB}$. Using the SVD of $\mathbf{S}$, it follows that

$$\mathbf{AU}\begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^\mathsf{T} = \mathbf{U}\begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^\mathsf{T}\mathbf{B}. \tag{A.1}$$

Since $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, left and right multiplying (A.1) by $\mathbf{U}^\mathsf{T}$ and $\mathbf{V}$, respectively, yields

$$\mathbf{U}^\mathsf{T}\mathbf{AU}\begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^\mathsf{T}\mathbf{BV}. \tag{A.2}$$

Define

$$\mathbf{U}^\mathsf{T}\mathbf{AU} = \begin{bmatrix} \bar{\mathbf{A}}_{11} & \bar{\mathbf{A}}_{12} \\ \bar{\mathbf{A}}_{21} & \bar{\mathbf{A}}_{22} \end{bmatrix}, \qquad \mathbf{V}^\mathsf{T}\mathbf{BV} = \begin{bmatrix} \bar{\mathbf{B}}_{11} & \bar{\mathbf{B}}_{12} \\ \bar{\mathbf{B}}_{21} & \bar{\mathbf{B}}_{22} \end{bmatrix}. \tag{A.3}$$

Substituting (A.3) into (A.2) results in

$$\begin{bmatrix} \bar{\mathbf{A}}_{11}\boldsymbol{\Sigma}_1 & \mathbf{0} \\ \bar{\mathbf{A}}_{21}\boldsymbol{\Sigma}_1 & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_1\bar{\mathbf{B}}_{11} & \boldsymbol{\Sigma}_1\bar{\mathbf{B}}_{12} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{A.4}$$

Equating similar terms in (A.4) yields $\bar{\mathbf{A}}_{11}\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1\bar{\mathbf{B}}_{11}$, $\bar{\mathbf{A}}_{21}\boldsymbol{\Sigma}_1 = \mathbf{0}$, and $\boldsymbol{\Sigma}_1\bar{\mathbf{B}}_{12} = \mathbf{0}$. Since $\boldsymbol{\Sigma}_1$ is full rank, it follows that

$$\bar{\mathbf{A}}_{11} = \boldsymbol{\Sigma}_1\bar{\mathbf{B}}_{11}\boldsymbol{\Sigma}_1^{-1}, \qquad \bar{\mathbf{A}}_{21} = \mathbf{0}, \qquad \bar{\mathbf{B}}_{12} = \mathbf{0}. \tag{A.5}$$

Substituting (A.5) into (A.3) and rearranging recovers the desired result. $\square$

Appendix B

Eigenvalues of Kronecker Products

Consider an LTI system with the minimal state-space realization

$$\boldsymbol{\mathcal{G}}_0 \overset{\min}{\sim} \left[\begin{array}{c|c} \mathbf{A}_0 & \mathbf{B}_0 \\ \hline \mathbf{C}_0 & \mathbf{D}_0 \end{array}\right],$$

with $\mathbf{A}_0, \mathbf{B}_0, \mathbf{C}_0, \mathbf{D}_0$ being matrices of appropriate dimensions. Assume $\exists \mathbf{P}_0 = \mathbf{P}_0^\mathsf{T} \succ 0$ such that

$$\mathbf{M}_0 = \begin{bmatrix} \mathbf{A}_0^\mathsf{T}\mathbf{P}_0\mathbf{A}_0 - \mathbf{P}_0 & \mathbf{A}_0^\mathsf{T}\mathbf{P}_0\mathbf{B}_0 - \mathbf{C}_0^\mathsf{T} \\ \left(\mathbf{A}_0^\mathsf{T}\mathbf{P}_0\mathbf{B}_0 - \mathbf{C}_0^\mathsf{T}\right)^\mathsf{T} & \mathbf{B}_0^\mathsf{T}\mathbf{P}_0\mathbf{B}_0 - \left(\mathbf{D}_0 + \mathbf{D}_0^\mathsf{T}\right) \end{bmatrix} \preceq 0. \tag{B.1}$$

Then, as per Lemma 3.1, $\boldsymbol{\mathcal{G}}_0$ is positive real. Now, consider the LTI system with the minimal state-space realization

$$\boldsymbol{\mathcal{G}} \overset{\min}{\sim} \left[\begin{array}{c|c} \mathbf{A}_0 \otimes \mathbf{1} & \mathbf{B}_0 \otimes \mathbf{1} \\ \hline \mathbf{C}_0 \otimes \mathbf{1} & \mathbf{D}_0 \otimes \mathbf{1} \end{array}\right].$$

As per [51, Theorem 4.2.15], $\mathbf{M} = \mathbf{M}_0 \otimes \mathbf{1} \preceq 0$ and $\mathbf{P} = \mathbf{P}_0 \otimes \mathbf{1} \succ 0$. As such, if $\boldsymbol{\mathcal{G}}_0$ is positive real, then $\boldsymbol{\mathcal{G}}$ is also positive real.

Appendix C

Proofs of Matrix-Scheduling of QSR-Dissipative Systems Theorems

Proof of Lemma 6.1. Applying the Rayleigh inequality to the QSR-dissipativity property in (6.5) yields

$$\tilde{V}_i \leq \lambda_{\max}(\mathbf{Q}_i)\|\mathbf{y}_i\|_{2T}^2 + 2\left\langle \mathbf{y}_i, \mathbf{S}_i\mathbf{u}_i \right\rangle_T + \lambda_{\max}(\mathbf{R}_i)\|\mathbf{u}_i\|_{2T}^2, \tag{C.1}$$

where $\lambda_{\max}(\mathbf{Q}_i) \in \mathbb{R}_{<0}$ and $\lambda_{\max}(\mathbf{R}_i) \in \mathbb{R}$ since $\mathbf{Q}_i \in \mathbb{S}_{--}^{n_y}$ and $\mathbf{R}_i \in \mathbb{S}^{n_y}$, respectively. Moreover, based on the pseudo-commutativity condition in Definition 6.3, it follows that $\boldsymbol{\Phi}_{y,i}^{\mathsf{T}}(t)\mathbf{S}_i = \mathbf{S}_i\boldsymbol{\Phi}_{u,i}(t)$ for all $t \in \mathbb{Z}_{\geq 0}$. Consider the term $\left\langle \mathbf{y}_i, \mathbf{S}_i\mathbf{u}_i \right\rangle_T$ in (C.1). Using the relation in (6.3), it follows that

$$\left\langle \mathbf{y}_i, \mathbf{S}_i\mathbf{u}_i \right\rangle_T = \left\langle \mathbf{y}_i, \mathbf{S}_i\boldsymbol{\Phi}_{u,i}\mathbf{u} \right\rangle_T = \left\langle \mathbf{y}_i, \boldsymbol{\Phi}_{y,i}^{\mathsf{T}}\mathbf{S}_i\mathbf{u} \right\rangle_T = \left\langle \boldsymbol{\Phi}_{y,i}\mathbf{y}_i, \mathbf{S}_i\mathbf{u} \right\rangle_T = \left\langle \bar{\mathbf{y}}_i, \mathbf{S}_i\mathbf{u} \right\rangle_T. \tag{C.2}$$

Additionally, define $\sigma_{y,i}(t)$ as the largest singular value of $\boldsymbol{\Phi}_{y,i}(t)$. From (6.3b), it follows that

$$\|\bar{\mathbf{y}}_i\|_{2T}^2 = \left\|\boldsymbol{\Phi}_{y,i}\mathbf{y}_i\right\|_{2T}^2 \leq \sum_{t\in\mathcal{T}} \sigma_{y,i}^2(t)\|\mathbf{y}_i(t)\|_2^2 \leq \bar{\sigma}_{y,i}^2\|\mathbf{y}_i\|_{2T}^2, \tag{C.3}$$

where

$$\bar{\sigma}_{y,i} = \sup_{t\in\mathbb{Z}_{\geq 0}} \sigma_{y,i}(t). \tag{C.4}$$

Furthermore, $\bar{\sigma}_{y,i} \in \mathbb{R}_{>0}$, provided there exists a $t \in \mathbb{Z}_{\geq 0}$ such that $\boldsymbol{\Phi}_{y,i}(t) \neq \mathbf{0}$. This condition is satisfied by design, since if this subsystem is being included in the parallel interconnection, it means at some point in time, its output must be nonzero, and therefore, its output scheduling matrix cannot be zero for all time. Additionally, $\bar{\sigma}_{y,i} < \infty$, since as per (6.7), the scheduling matrices are assumed to be bounded. Consequently, rearranging (C.3) yields

$$\frac{1}{\bar{\sigma}_{y,i}^2}\|\bar{\mathbf{y}}_i\|_{2T}^2 \leq \|\mathbf{y}_i\|_{2T}^2. \tag{C.5}$$

Since $\lambda_{\max}(\mathbf{Q}_i) \in \mathbb{R}_{<0}$, substituting (C.2) and (C.5) into (C.1) leads to

$$\tilde{V}_i \leq \frac{\lambda_{\max}(\mathbf{Q}_i)}{\bar{\sigma}_{y,i}^2} \|\bar{\mathbf{y}}_i\|_{2T}^2 + 2\left\langle \bar{\mathbf{y}}_i, \mathbf{S}_i \mathbf{u} \right\rangle_T + \lambda_{\max}(\mathbf{R}_i) \|\mathbf{u}_i\|_{2T}^2. \tag{C.6}$$

Given that $\bar{\sigma}_{y,i}^2 \in \mathbb{R}_{>0}$, (C.6) can be written as

$$\bar{\sigma}_{y,i}^2 \tilde{V}_i \leq -\varepsilon_i \|\bar{\mathbf{y}}_i\|_{2T}^2 + 2\bar{\sigma}_{y,i}^2 \left\langle \bar{\mathbf{y}}_i, \mathbf{S}_i \mathbf{u} \right\rangle_T + \lambda_{\max}(\mathbf{R}_i) \bar{\sigma}_{y,i}^2 \|\mathbf{u}_i\|_{2T}^2, \tag{C.7}$$

where

$$\varepsilon_i = |\lambda_{\max}(\mathbf{Q}_i)| = -\lambda_{\max}(\mathbf{Q}_i) > 0. \tag{C.8}$$

Since other than symmetry, there are no restrictions on $\mathbf{R}_i$, the sign of its maximum eigenvalue is unknown. Denoting $\sigma_{u,i}(t)$ and $\nu_{u,i}(t)$ as the largest and smallest singular values of $\mathbf{\Phi}_{u,i}(t)$, respectively, define

$$\delta_i = \begin{cases} \lambda_{\max}(\mathbf{R}_i) \bar{\sigma}_{y,i}^2 \bar{\sigma}_{u,i}^2, & \text{if } \lambda_{\max}(\mathbf{R}_i) > 0, \\ \lambda_{\max}(\mathbf{R}_i) \bar{\sigma}_{y,i}^2 \bar{\nu}_{u,i}^2, & \text{if } \lambda_{\max}(\mathbf{R}_i) \leq 0, \end{cases} \tag{C.9}$$

where

$$\bar{\sigma}_{u,i} = \sup_{t \in \mathbb{Z}_{\geq 0}} \sigma_{u,i}(t), \qquad \bar{\nu}_{u,i} = \inf_{t \in \mathbb{Z}_{\geq 0}} \nu_{u,i}(t). \tag{C.10}$$

If $\mathbf{S}_i$ is full rank and $n_u \leq n_y$, the pseudo-commutativity property suggests $\bar{\sigma}_{u,i} \in \mathbb{R}_{>0}$, since $\bar{\sigma}_{y,i} \in \mathbb{R}_{>0}$. If $\mathbf{S}_i$ is rank deficient or $n_y < n_u$, it is possible to have $\bar{\sigma}_{u,i} \in \mathbb{R}_{\geq 0}$, since the output scheduling matrices can be designed with $\mathbf{\Phi}_{u,i}(t) = \mathbf{0}$ for all $t \in \mathbb{Z}_{\geq 0}$. Additionally, $\bar{\nu}_{u,i} \in \mathbb{R}_{\geq 0}$, since the scheduling matrix $\mathbf{\Phi}_{u,i}(t)$ may not be full rank at all times. Consider the term $\lambda_{\max}(\mathbf{R}_i) \bar{\sigma}_{y,i}^2 \|\mathbf{u}_i\|_{2T}^2$ in (C.7). Using (6.3a) and the definition of δ_i in (C.9), it follows that

$$\lambda_{\max}(\mathbf{R}_i) \bar{\sigma}_{y,i}^2 \|\mathbf{u}_i\|_{2T}^2 = \lambda_{\max}(\mathbf{R}_i) \bar{\sigma}_{y,i}^2 \left\| \mathbf{\Phi}_{u,i} \mathbf{u} \right\|_{2T}^2 \leq \delta_i \|\mathbf{u}\|_{2T}^2. \tag{C.11}$$

Substituting (C.11) into (C.7) yields

$$\begin{aligned} \hat{V}_i &\leq -\varepsilon_i \|\bar{\mathbf{y}}_i\|_{2T}^2 + 2\bar{\sigma}_{y,i}^2 \left\langle \bar{\mathbf{y}}_i, \mathbf{S}_i \mathbf{u} \right\rangle_T + \delta_i \|\mathbf{u}\|_{2T}^2 \\ &= \left\langle \bar{\mathbf{y}}_i, \bar{\mathbf{Q}}_i \bar{\mathbf{y}}_i \right\rangle_T + 2\left\langle \bar{\mathbf{y}}_i, \bar{\mathbf{S}}_i \mathbf{u} \right\rangle_T + \left\langle \mathbf{u}, \bar{\mathbf{R}}_i \mathbf{u} \right\rangle_T, \end{aligned} \tag{C.12}$$

where $\hat{V}_i = \bar{\sigma}_{y,i}^2 \tilde{V}_i$, and

$$\bar{\mathbf{Q}}_i = -\varepsilon_i \mathbf{1}, \qquad \bar{\mathbf{S}}_i = \bar{\sigma}_{y,i}^2 \mathbf{S}_i, \qquad \bar{\mathbf{R}}_i = \delta_i \mathbf{1}. \tag{C.13}$$

Consequently, the gain-scheduled subsystem $\bar{\boldsymbol{\mathcal{G}}}_i$ is QSR-dissipative with $\bar{\mathbf{Q}}_i \in \mathbb{S}_{--}^{n_y}$, $\bar{\mathbf{S}}_i \in \mathbb{R}^{n_y \times n_u}$, and $\bar{\mathbf{R}}_i \in \mathbb{S}^{n_u}$ as defined in (C.13). □

Proof of Theorem 6.1. Following Lemma 6.1, each scheduled subsystem $\bar{\boldsymbol{\mathcal{G}}}_i$ is QSR-dissipative with $\bar{\mathbf{Q}}_i \in \mathbb{S}_{--}^{n_y}$, $\bar{\mathbf{S}}_i \in \mathbb{R}^{n_y \times n_u}$, and $\bar{\mathbf{R}}_i \in \mathbb{S}^{n_u}$ defined in (C.13). Furthermore, $\bar{\sigma}_{y,i}$, ε_i, and δ_i are given by (C.4), (C.8), and (C.9), respectively. Adding and subtracting $\bar{\sigma}_{y,i}^4/\varepsilon_i \|\mathbf{S}_i \mathbf{u}\|_{2T}^2$ from (C.12) and factoring leads to

$$\hat{V}_i \leq -\varepsilon_i \left\| \bar{\mathbf{y}}_i - \frac{\bar{\sigma}_{y,i}^2}{\varepsilon_i} \mathbf{S}_i \mathbf{u} \right\|_{2T}^2 + \delta_i \|\mathbf{u}\|_{2T}^2 + \frac{\bar{\sigma}_{y,i}^4}{\varepsilon_i} \|\mathbf{S}_i \mathbf{u}\|_{2T}^2. \tag{C.14}$$

To simplify further analysis, let

$$\tilde{\mathbf{y}}_i(t) = \bar{\mathbf{y}}_i(t) - \frac{\bar{\sigma}_{y,i}^2}{\varepsilon_i} \mathbf{S}_i \mathbf{u}(t), \tag{C.15}$$

for $i \in \mathcal{N}$ and denote $\sigma_{\mathbf{S}_i}$ as the largest singular value of $\mathbf{S}_i$. Substituting (C.15) into (C.14) and using the Rayleigh inequality results in

$$\hat{V}_i \leq -\varepsilon_i \|\tilde{\mathbf{y}}_i\|_{2T}^2 + \left(\delta_i + \frac{\bar{\sigma}_{y,i}^4 \sigma_{\mathbf{S}_i}^2}{\varepsilon_i} \right) \|\mathbf{u}\|_{2T}^2. \tag{C.16}$$

Summing (C.16) over all $i \in \mathcal{N}$, it follows that

$$\begin{aligned} \sum_{i \in \mathcal{N}} \hat{V}_i &\leq -\sum_{i \in \mathcal{N}} \varepsilon_i \|\tilde{\mathbf{y}}_i\|_{2T}^2 + \sum_{i \in \mathcal{N}} \left(\delta_i + \frac{\bar{\sigma}_{y,i}^4 \sigma_{\mathbf{S}_i}^2}{\varepsilon_i} \right) \|\mathbf{u}\|_{2T}^2 \\ &\leq -\varepsilon_{\min} \sum_{i \in \mathcal{N}} \|\tilde{\mathbf{y}}_i\|_{2T}^2 + \bar{\delta} \|\mathbf{u}\|_{2T}^2, \end{aligned} \tag{C.17}$$

where

$$\varepsilon_{\min} = \min_{i \in \mathcal{N}} \varepsilon_i > 0, \qquad \bar{\delta} = \sum_{i \in \mathcal{N}} \left(\delta_i + \frac{\bar{\sigma}_{y,i}^4 \sigma_{\mathbf{S}_i}^2}{\varepsilon_i} \right). \tag{C.18}$$

Rewriting the summation in (C.17) as a vector norm yields

$$\sum_{i\in\mathcal{N}} \hat{V}_i \leq -\varepsilon_{\min} \left\| \begin{bmatrix} \|\tilde{\mathbf{y}}_1\|_{2T} \\ \vdots \\ \|\tilde{\mathbf{y}}_N\|_{2T} \end{bmatrix} \right\|_2^2 + \bar{\delta}\|\mathbf{u}\|_{2T}^2. \tag{C.19}$$

Applying the inequality in Lemma 2.3, which is a special case of the Cauchy–Schwartz inequality, to (C.19) results in

$$\begin{aligned} \sum_{i\in\mathcal{N}} \hat{V}_i &\leq -\frac{\varepsilon_{\min}}{N} \left\langle \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} \|\tilde{\mathbf{y}}_1\|_{2T} \\ \vdots \\ \|\tilde{\mathbf{y}}_N\|_{2T} \end{bmatrix} \right\rangle^2 + \bar{\delta}\|\mathbf{u}\|_{2T}^2 \\ &= -\frac{\varepsilon_{\min}}{N} \left(\sum_{i\in\mathcal{N}} \|\tilde{\mathbf{y}}_i\|_{2T} \right)^2 + \bar{\delta}\|\mathbf{u}\|_{2T}^2. \end{aligned} \tag{C.20}$$

Multiplying both sides of (C.20) by N and applying the triangle inequality yields

$$N\sum_{i\in\mathcal{N}} \hat{V}_i \leq -\varepsilon_{\min} \left\| \sum_{i\in\mathcal{N}} \tilde{\mathbf{y}}_i \right\|_{2T}^2 + N\bar{\delta}\|\mathbf{u}\|_{2T}^2. \tag{C.21}$$

Substituting back (C.15) into (C.21) and defining $\tilde{V} = N\sum_{i\in\mathcal{N}} \hat{V}_i$ results in

$$\tilde{V} \leq -\varepsilon_{\min} \left\| \sum_{i\in\mathcal{N}} \left(\bar{\mathbf{y}}_i - \frac{\bar{\sigma}_{y,i}^2}{\varepsilon_i} \mathbf{S}_i \mathbf{u} \right) \right\|_{2T}^2 + N\bar{\delta}\|\mathbf{u}\|_{2T}^2. \tag{C.22}$$

Defining $\bar{\mathbf{S}} = \sum_{i\in\mathcal{N}} \left(\bar{\sigma}_{y,i}^2/\varepsilon_i\right)\mathbf{S}_i$ and using (6.4), it follows from (C.22) that

$$\begin{aligned} \tilde{V} &\leq -\varepsilon_{\min} \left\|\mathbf{y} - \bar{\mathbf{S}}\mathbf{u}\right\|_{2T}^2 + N\bar{\delta}\|\mathbf{u}\|_{2T}^2 \\ &= -\varepsilon_{\min} \left(\|\mathbf{y}\|_{2T}^2 - 2\left\langle \mathbf{y}, \bar{\mathbf{S}}\mathbf{u} \right\rangle_T + \left\|\bar{\mathbf{S}}\mathbf{u}\right\|_{2T}^2 \right) + N\bar{\delta}\|\mathbf{u}\|_{2T}^2. \end{aligned} \tag{C.23}$$

Applying the Rayleigh inequality to (C.23) and defining $\nu_{\bar{\mathbf{S}}}$ as the smallest singular value of $\bar{\mathbf{S}}$, it follows that

$$\begin{aligned} \tilde{V} &\leq -\varepsilon_{\min}\|\mathbf{y}\|_{2T}^2 + 2\varepsilon_{\min} \left\langle \mathbf{y}, \bar{\mathbf{S}}\mathbf{u} \right\rangle_T + \left(N\bar{\delta} - \varepsilon_{\min}\nu_{\bar{\mathbf{S}}}^2 \right)\|\mathbf{u}\|_{2T}^2 \\ &= -\varepsilon_{\min}\|\mathbf{y}\|_{2T}^2 + 2\varepsilon_{\min} \left\langle \mathbf{y}, \bar{\mathbf{S}}\mathbf{u} \right\rangle_T + \hat{\delta}\|\mathbf{u}\|_{2T}^2, \end{aligned} \tag{C.24}$$

where $\hat{\delta} = N\bar{\delta} - \varepsilon_{\min}\nu_{\bar{\mathbf{S}}}^2$. Alternatively, (C.24) can be written as

$$\tilde{V} \leq \langle \mathbf{y}, \mathbf{Q}\mathbf{y} \rangle_T + 2\langle \mathbf{y}, \mathbf{S}\mathbf{u} \rangle_T + \langle \mathbf{u}, \mathbf{R}\mathbf{u} \rangle_T,$$

where

$$\mathbf{Q} = -\varepsilon_{\min}\mathbf{1}, \qquad \mathbf{S} = \varepsilon_{\min}\bar{\mathbf{S}}, \qquad \mathbf{R} = \hat{\delta}\mathbf{1}. \tag{C.25}$$

Additionally, (C.25) can be written in terms of $\bar{\sigma}_{y,i}$, ε_i, and δ_i, given by (C.4), (C.8), and (C.9), respectively, as follows:

$$\mathbf{Q} = -\min_{i\in\mathcal{N}}(\varepsilon_i)\mathbf{1}, \tag{C.26a}$$

$$\mathbf{S} = \min_{i\in\mathcal{N}}(\varepsilon_i)\sum_{i\in\mathcal{N}}\frac{\bar{\sigma}_{y,i}^2}{\varepsilon_i}\mathbf{S}_i, \tag{C.26b}$$

$$\mathbf{R} = \left(N\sum_{i\in\mathcal{N}}\left(\delta_i + \frac{\bar{\sigma}_{y,i}^4\sigma_{\mathbf{S}_i}^2}{\varepsilon_i}\right) - \min_{i\in\mathcal{N}}(\varepsilon_i)\nu_{\bar{\mathbf{S}}}^2\right)\mathbf{1}, \tag{C.26c}$$

where $\sigma_{\mathbf{S}_i}$ is the largest singular value of $\mathbf{S}_i$ and $\nu_{\bar{\mathbf{S}}}$ is the smallest singular value of $\bar{\mathbf{S}} = \sum_{i\in\mathcal{N}}\big(\bar{\sigma}_{y,i}^2/\varepsilon_i\big)\mathbf{S}_i$. Therefore, the gain-scheduled system $\bar{\boldsymbol{\mathcal{G}}}$ is QSR-dissipative with $\mathbf{Q} \in \mathbb{S}_{--}^{n_y}$, $\mathbf{S} \in \mathbb{R}^{n_y\times n_u}$, and $\mathbf{R} \in \mathbb{S}^{n_u}$ as defined in (C.26). □

Proof of Lemma 6.2. Using the Rayleigh inequality, it follows that

$$\sum_{i\in\mathcal{N}}\langle \mathbf{y}_i, \mathbf{Q}_i\mathbf{y}_i \rangle_T \leq -\sum_{i\in\mathcal{N}}\varepsilon_i\|\mathbf{y}_i\|_{2T}^2 \leq -\varepsilon_{\min}\sum_{i\in\mathcal{N}}\|\mathbf{y}_i\|_{2T}^2, \tag{C.27}$$

where $\varepsilon_i = -\lambda_{\max}(\mathbf{Q}_i) \geq 0$ and $\varepsilon_{\min} = \min_{i\in\mathcal{N}}\varepsilon_i \geq 0$. Additionally, the summation in (6.4) can be rewritten as the matrix-vector multiplication $\mathbf{y}(t) = \boldsymbol{\Psi}(t)\boldsymbol{\upsilon}(t)$, where

$$\boldsymbol{\Psi}(t) = \begin{bmatrix}\boldsymbol{\Phi}_{y,1}(t) & \cdots & \boldsymbol{\Phi}_{y,N}(t)\end{bmatrix}, \qquad \boldsymbol{\upsilon}(t) = \begin{bmatrix}\mathbf{y}_1(t)\\ \vdots \\ \mathbf{y}_N(t)\end{bmatrix}. \tag{C.28}$$

Consequently, by defining $\sigma_{\boldsymbol{\Psi}}(t)$ as the largest singular value of $\boldsymbol{\Psi}(t)$ and applying the Rayleigh inequality to $\|\boldsymbol{\Psi}(t)\boldsymbol{\upsilon}(t)\|_2^2$, it follows that

$$\|\mathbf{y}(t)\|_2^2 = \|\boldsymbol{\Psi}(t)\boldsymbol{\upsilon}(t)\|_2^2 \leq \sigma_{\boldsymbol{\Psi}}^2(t)\|\boldsymbol{\upsilon}(t)\|_2^2. \tag{C.29}$$

Moreover, $\sigma_{\boldsymbol{\Psi}}(t) \in \mathbb{R}_{>0}$, provided the output scheduling matrices are active, that is, $\forall t \in \mathbb{Z}_{\geq 0}$,

$\exists i \in \mathcal{N}$ such that $\mathbf{\Phi}_{y,i}(t) \neq \mathbf{0}$, and therefore, $\mathbf{\Psi}^\mathsf{T}(t)\mathbf{\Psi}(t) \in \mathbb{S}_{+}^{N\times n_y}$ is nonzero. Additionally,

$$\sigma_{\mathbf{\Psi}}^2(t) \leq \operatorname{tr}\bigl(\mathbf{\Psi}^\mathsf{T}(t)\mathbf{\Psi}(t)\bigr) = \sum_{i\in\mathcal{N}} \operatorname{tr}\bigl(\mathbf{\Phi}_{y,i}^\mathsf{T}(t)\mathbf{\Phi}_{y,i}(t)\bigr) \leq n_y \sum_{i\in\mathcal{N}} \lambda_{\max}\bigl(\mathbf{\Phi}_{y,i}^\mathsf{T}(t)\mathbf{\Phi}_{y,i}(t)\bigr) = n_y \sum_{i\in\mathcal{N}} \sigma_{y,i}^2(t).$$

Therefore, $\sup_{t\in\mathbb{Z}_{\geq 0}} \sigma_{\mathbf{\Psi}}^2(t) < \infty$ since the output scheduling matrices are assumed to be bounded as per (6.7). Consequently, rearranging (C.29) yields

$$\frac{1}{\sigma_{\mathbf{\Psi}}^2(t)}\|\mathbf{y}(t)\|_2^2 \leq \|\boldsymbol{\upsilon}(t)\|_2^2 = \sum_{i\in\mathcal{N}}\|\mathbf{y}_i(t)\|_2^2. \tag{C.30}$$

Since $\lambda_{\max}(\mathbf{Q}_i) \in \mathbb{R}_{\leq 0}$, substituting (C.30) into (C.27), it follows that

$$\sum_{i\in\mathcal{N}} \langle \mathbf{y}_i, \mathbf{Q}_i\mathbf{y}_i\rangle_T \leq -\varepsilon_{\min}\sum_{t\in\mathcal{T}} \frac{1}{\sigma_{\mathbf{\Psi}}^2(t)}\|\mathbf{y}(t)\|_2^2 \leq -\frac{\varepsilon_{\min}}{\bar{\sigma}_{\mathbf{\Psi}}^2}\sum_{t\in\mathcal{T}}\|\mathbf{y}(t)\|_2^2 = -\varepsilon\|\mathbf{y}\|_{2T}^2 = \langle \mathbf{y}, \mathbf{Q}\mathbf{y}\rangle_T\,,$$

where

$$\bar{\sigma}_{\mathbf{\Psi}} = \sup_{t\in\mathbb{Z}_{\geq 0}} \sigma_{\mathbf{\Psi}}(t) > 0, \qquad \varepsilon = \frac{\varepsilon_{\min}}{\bar{\sigma}_{\mathbf{\Psi}}^2} \geq 0, \tag{C.31}$$

and $\mathbf{Q} = -\varepsilon\mathbf{1} \in \mathbb{S}_{-}^{n_y}$. □

Proof of Lemma 6.3. Using the Rayleigh inequality, it follows that

$$\sum_{i\in\mathcal{N}} \langle \mathbf{u}_i, \mathbf{R}_i\mathbf{u}_i\rangle_T \leq \sum_{i\in\mathcal{N}} \lambda_{\max}(\mathbf{R}_i)\|\mathbf{u}_i\|_{2T}^2. \tag{C.32}$$

Given the index sets (6.12), it follows that

$$\sum_{i\in\mathcal{N}} \lambda_{\max}(\mathbf{R}_i) = \sum_{i\in\mathcal{R}_{>0}} \lambda_{\max}(\mathbf{R}_i) + \sum_{i\in\mathcal{R}_{<0}} \lambda_{\max}(\mathbf{R}_i). \tag{C.33}$$

Expanding (C.32) using (C.33) results in

$$\sum_{i\in\mathcal{N}} \langle \mathbf{u}_i, \mathbf{R}_i\mathbf{u}_i\rangle_T \leq \sum_{i\in\mathcal{R}_{>0}} \lambda_{\max}(\mathbf{R}_i)\|\mathbf{u}_i\|_{2T}^2 + \sum_{i\in\mathcal{R}_{<0}} \lambda_{\max}(\mathbf{R}_i)\|\mathbf{u}_i\|_{2T}^2. \tag{C.34}$$

Substituting the relation in (6.3a) into (C.34), it follows that

$$\sum_{i\in\mathcal{N}} \langle \mathbf{u}_i, \mathbf{R}_i\mathbf{u}_i\rangle_T \leq \delta_{\max} \sum_{i\in\mathcal{R}_{>0}} \left\|\mathbf{\Phi}_{u,i}\mathbf{u}\right\|_{2T}^2 - \delta_{\min} \sum_{i\in\mathcal{R}_{<0}} \left\|\mathbf{\Phi}_{u,i}\mathbf{u}\right\|_{2T}^2, \tag{C.35}$$

where

$$\delta_i = |\lambda_{\max}(\mathbf{R}_i)|, \qquad \delta_{\max} = \max_{i \in \mathcal{R}_{>0}} \delta_i, \qquad \delta_{\min} = \min_{i \in \mathcal{R}_{<0}} \delta_i.$$

Expanding (C.35) and using the Rayleigh inequality yields

$$\begin{aligned}
\sum_{i \in \mathcal{N}} \langle \mathbf{u}_i, \mathbf{R}_i \mathbf{u}_i \rangle_T &\leq \delta_{\max} \sum_{i \in \mathcal{R}_{>0}} \sum_{t \in \mathcal{T}} \sigma_{u,i}^2(t) \|\mathbf{u}(t)\|_2^2 \; - \delta_{\min} \sum_{i \in \mathcal{R}_{<0}} \sum_{t \in \mathcal{T}} \nu_{u,i}^2(t) \|\mathbf{u}(t)\|_2^2 \\
&\leq \delta_{\max} \bar{\sigma}_u \sum_{t \in \mathcal{T}} \|\mathbf{u}(t)\|_2^2 - \delta_{\min} \bar{\nu}_u \sum_{t \in \mathcal{T}} \|\mathbf{u}(t)\|_2^2 = \delta \|\mathbf{u}\|_{2T}^2 = \langle \mathbf{u}, \mathbf{R}\mathbf{u} \rangle_T \,,
\end{aligned}$$

where

$$\bar{\sigma}_u = \sup_{t \in \mathbb{Z}_{\geq 0}} \sum_{i \in \mathcal{R}_{>0}} \sigma_{u,i}^2(t) \geq 0, \tag{C.36a}$$

$$\bar{\nu}_u = \inf_{t \in \mathbb{Z}_{\geq 0}} \sum_{i \in \mathcal{R}_{<0}} \nu_{u,i}^2(t) \geq 0, \tag{C.36b}$$

$$\delta = \delta_{\max} \bar{\sigma}_u - \delta_{\min} \bar{\nu}_u, \tag{C.36c}$$

and $\mathbf{R} = \delta \mathbf{1}$. $\square$

Bibliography

Bibliography

[1] S. Moalemi and J. R. Forbes, "Input-Output Stability of Gradient Descent: A Discrete-Time Passivity-Based Approach", *American Control Conference*, pp. 924–929, 2025. DOI: 10.23919/ACC63710.2025.11107445.

[2] C. J. Damaren, "Gain-Scheduled SPR Controllers for Nonlinear Flexible Systems", *Journal of Dynamic Systems, Measurement, and Control*, vol. 118, no. 4, pp. 698–703, 1996. DOI: 10.1115/1.2802345.

[3] S. Moalemi and J. R. Forbes, "Passivity-Based Gain-Scheduled Control with Scheduling Matrices", *IEEE Conference on Control Technology and Applications*, pp. 7–13, 2024. DOI: 10.1109/CCTA60707.2024.10666540.

[4] S. Moalemi and J. R. Forbes, "Matrix-Scheduling of QSR-Dissipative Systems", *IEEE Transactions on Automatic Control*, vol. 70, no. 8, pp. 5286–5300, 2025. DOI: 10.1109/TAC.2025.3542329.

[5] B. Brogliato, R. Lozano, B. Maschke, and O. Egeland, *Dissipative Systems Analysis and Control: Theory and Applications*. Germany: Springer London, 2013. DOI: 10.1007/978-3-030-19420-8_4.

[6] B. T. Polyak, "Some Methods of Speeding Up the Convergence of Iteration Methods", *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964. DOI: 10.1016/0041-5553(64)90137-5.

[7] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*. USA: Springer, 2013. DOI: 10.1007/978-1-4419-8853-9.

[8] V. Ugrinovskii, I. R. Petersen, and I. Shames, "Global Convergence and Asymptotic Optimality of the Heavy Ball Method for a Class of Nonconvex Optimization Problems", *IEEE Control Systems Letters*, vol. 6, pp. 2449–2454, 2022. DOI: 10.1109/lcsys.2022.3163408.

[9] A. Wu, I. Petersen, V. Ugrinovskii, and I. Shames, "A Generalized Accelerated Gradient Optimization Method", *American Control Conference*, pp. 1904–1908, 2024. DOI: 10.23919/acc60939.2024.10644352.

[10] B. Hu and L. Lessard, "Control Interpretations for First-Order Optimization Methods", *American Control Conference*, pp. 3114–3119, 2017. DOI: 10.23919/acc.2017.7963426.

[11] I. Notarnicola and A. Falsone, "Passivity-based Analysis of the ADMM Algorithm for Constraint-Coupled Optimization", *Automatica*, vol. 146, p. 110 552, 2022. DOI: 10.1016/j.automatica.2022.110552.

[12] L. Lessard, "The Analysis of Optimization Algorithms: A Dissipativity Approach", *IEEE Control Systems Magazine*, vol. 42, no. 3, pp. 58–72, 2022. DOI: 10.1109/mcs.2022.3157115.

[13] B. Van Scoy and L. Lessard, "A Tutorial on a Lyapunov-Based Approach to the Analysis of Iterative Optimization Algorithms", *IEEE Conference on Decision and Control*, pp. 3003–3008, 2023. DOI: 10.1109/cdc49753.2023.10384074.

[14] 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. DOI: 10.1137/15m1009597.

[15] 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, 2017. DOI: 10.1109/lcsys.2017.2722406.

[16] R. Padmanabhan and P. Seiler, "Analysis of Gradient Descent with Varying Step Sizes using Integral Quadratic Constraints", *IEEE Transactions on Automatic Control*, vol. 70, no. 1, pp. 587–594, 2025. DOI: 10.1109/tac.2024.3438808.

[17] J. R. Forbes and C. J. Damaren, "Design of Gain-Scheduled Strictly Positive Real Controllers Using Numerical Optimization for Flexible Robotic Systems", *Journal of Dynamic Systems, Measurement, and Control*, vol. 132, no. 3, 2010. DOI: 10.1115/1.4001335.

[18] A. Walsh and J. R. Forbes, "A Very Strictly Passive Gain-Scheduled Controller: Theory and Experiments", *IEEE/ASME Transactions on Mechatronics*, vol. 21, no. 6, pp. 2817–2826, 2016. DOI: 10.1109/tmech.2016.2565387.

[19] A. Walsh and J. R. Forbes, "Very Strictly Passive Controller Synthesis With Affine Parameter Dependence", *IEEE Transactions on Automatic Control*, vol. 63, no. 5, pp. 1531–1537, 2018. DOI: 10.1109/tac.2017.2748928.

[20] J. J. Brown and R. Caverly, "Conic Gain-Scheduled Control of an Aeroelastic Airfoil", *AIAA Scitech 2021 Forum*, p. 1562, 2021. DOI: 10.2514/6.2021-1562.vid.

[21] L. Anderson, R. J. Caverly, and A. Lamperski, "Gain-Scheduled QSR-Dissipative Systems: An Input-Output Approach", *American Control Conference*, pp. 2417–2423, 2023. DOI: 10.23919/acc55779.2023.10156171.

[22] J. R. Forbes, "Extensions of Input-Output Stability Theory and the Control of Aerospace Systems", Ph.D. dissertation, University of Toronto, Canada, 2011.

[23] X. Lang and C. J. Damaren, "Gain-Scheduled Control for an Antenna with Multiple Collocated Sensors and Actuators", *AIAA Guidance, Navigation, and Control Conference*, pp. 1–22, 2018. DOI: 10.2514/6.2018-0858.

[24] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. United Kingdom: Prentice Hall, 1996.

[25] H. J. Márquez, *Nonlinear Control Systems: Analysis and Design*. United Kingdom: Wiley, 2003.

[26] D. Djukić, V. Janković, I. Matić, and N. Petrović, *The IMO Compendium: A Collection of Problems Suggested for The International Mathematical Olympiads: 1959-2009*, 2nd ed. Netherlands: Springer New York, 2011.

[27] F. R. Gantmacher, *The Theory of Matrices*. United Kingdom: Chelsea Publishing Company, 1959.

[28] C. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*. USA: Academic Press, 1975.

[29] P. Antsaklis and A. Michel, *Linear Systems*. Germany: McGraw-Hill, 2006.

[30] G. Zames, "On the Input-Output Stability of Time-Varying Nonlinear Feedback Systems Part I: Conditions Derived Using Concepts of Loop Gain, Conicity, and Positivity", *IEEE Transactions on Automatic Control*, vol. 11, no. 2, pp. 228–238, 1966. DOI: 10.1109/tac.1966.1098316.

[31] J. C. Willems, "Dissipative Dynamical Systems Part I: General Theory", *Archive for Rational Mechanics and Analysis*, vol. 45, pp. 321–351, 1972.

[32] L. Hitz and B. Anderson, "Discrete Positive-Real Functions and their Application to System Stability", *Proceedings of the Institution of Electrical Engineers*, vol. 116, no. 1, pp. 153–155, 1969.

[33] W. M. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. USA: Princeton University Press, 2008.

[34] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. United States: Cambridge University Press, 2004.

[35] B. Polyak, *Introduction to Optimization*. USA: Optimization Software, Inc., 1987.

[36] C. I. Byrnes and W. Lin, "Losslessness, Feedback Equivalence, and the Global Stabilization of Discrete-Time Nonlinear Systems", *IEEE Transactions on Automatic Control*, vol. 39, no. 1, pp. 83–98, 1994. DOI: 10.1109/9.273341.

[37] S. Kelly and J. W. Simpson-Porco, "An Interconnected Systems Approach to Convergence Analysis of Discrete-Time Primal-Dual Algorithms", *American Control Conference*, pp. 662–668, 2024. DOI: 10.23919/acc60939.2024.10644608.

[38] J. W. Simpson-Porco, "Equilibrium-Independent Dissipativity with Quadratic Supply Rates", *IEEE Transactions on Automatic Control*, vol. 64, no. 4, pp. 1440–1455, 2018. DOI: 10.1109/tac.2018.2838664.

[39] H. K. Khalil, *Nonlinear Systems*. United Kingdom: Prentice Hall, 1996.

[40] A. Van Der Schaft, *$\mathcal{L}_2$-Gain and Passivity Techniques in Nonlinear Control*. Germany: Springer International Publishing, 2018.

[41] D. Hill and P. Moylan, "The Stability of Nonlinear Dissipative Systems", *IEEE Transactions on Automatic Control*, vol. 21, no. 5, pp. 708–711, 1976. DOI: 10.1109/TAC.1976.1101352.

[42] J. L. Willems, "A System Theory Approach to Unified Electrical Machine Analysis", *International Journal of Control*, vol. 15, no. 3, pp. 401–418, 1972. DOI: 10.1080/00207177208932158.

[43] D. Hill and P. Moylan, "Stability Results for Nonlinear Feedback Systems", *Automatica*, vol. 13, no. 4, pp. 377–382, 1977. DOI: 10.1016/0005-1098(77)90020-6.

[44] A. L. Fradkov and D. J. Hill, "Exponential Feedback Passivity and Stabilizability of Nonlinear Systems", *Automatica*, vol. 34, no. 6, pp. 697–703, 1998. DOI: 10.1016/s0005-1098(97)00230-6.

[45] A. Rahnama, M. Xia, and P. J. Antsaklis, "A QSR-Dissipativity and Passivity Based Analysis of Event-Triggered Networked Control Systems", *IEEE Conference on Decision and Control*, pp. 3072–3077, 2016. DOI: 10.1109/cdc.2016.7798729.

[46] X. Zhang, Y. Liu, H. Qiu, and H. Liu, "Dissipativity and Synchronization of Fractional-Order Output-Coupled Neural Networks with Multiple Adaptive Coupling Weights", *Mathematics and Computers in Simulation*, vol. 215, pp. 306–322, 2024. DOI: 10.1016/j.matcom.2023.08.016.

[47] R. Ortega, A. Loría, P. J. Nicklasson, and H. Sira-Ramírez, *Passivity-based Control of Euler-Lagrange Systems: Mechanical, Electrical and Electromechanical Applications*. Germany: Springer London, 2013. DOI: 10.1108/ir.1999.04926cae.003.

[48] C. J. Damaren, “Passivity Analysis for Flexible Multilink Space Manipulators”, *Journal of Guidance Control and Dynamics*, vol. 18, no. 2, pp. 272–279, 1995. DOI: 10.2514/3.21380.

[49] M. Panzirsch, J. Artigas, J.-H. Ryu, and M. Ferre, “Multilateral control for delayed teleoperation”, *International Conference on Advanced Robotics*, pp. 1–6, 2013.

[50] J. R. Forbes and C. J. Damaren, “Overcoming Passivity Violations: Closed-Loop Stability, Controller Design and Controller Scheduling”, *IET Control Theory & Applications*, vol. 7, no. 6, pp. 785–795, 2013. DOI: 10.1049/iet-cta.2012.0459.

[51] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. United Kingdom: Cambridge University Press, 1994.

[52] L. Armijo, “Minimization of Functions having Lipschitz Continuous First Partial Derivatives”, *Pacific Journal of Mathematics*, vol. 16, no. 1, pp. 1–3, 1966. DOI: 10.2140/pjm.1966.16.1.