KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Performance Analysis of Positive Systems and Optimization Algorithms with Time-delays

HAMID REZA FEYZMAHDAVIAN

Doctoral Thesis
Stockholm, Sweden 2016

TRITA-EE 2015:83
ISSN 1653-5146
ISBN 978-91-7595-790-6

KTH Royal Institute of Technology
School of Electrical Engineering
Department of Automatic Control
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstrand av Kungliga Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i reglerteknik fredagen den 15 January 2016, klockan 10:00, i sal Q2, Kungliga Tekniska högskolan, Osquldas väg 10, Stockholm.

Tryck: Universitetsservice US AB

Abstract

Time-delay dynamical systems are used to model many real-world engineering systems, where the future evolution of a system depends not only on current states but also on the history of states. For this reason, the study of stability and control of time-delay systems is of theoretical and practical importance. In this thesis, we develop several stability analysis frameworks for dynamical systems in the presence of communication and computation time-delays, and apply our results to different challenging engineering problems.

The thesis first considers delay-independent stability of positive monotone systems. We show that the asymptotic stability of positive monotone systems whose vector fields are homogeneous is independent of the magnitude and variation of time-varying delays. We present explicit expressions that allow us to give explicit estimates of the decay rate for various classes of time-varying delays. For positive linear systems, we demonstrate that the best decay rate that our results guarantee can be found via convex optimization. We also derive a set of necessary and sufficient conditions for asymptotic stability of general positive monotone (not necessarily homogeneous) systems with time-delays. As an application of our theoretical results, we discuss delay-independent stability of continuous-time power control algorithms in wireless networks.

The thesis continues by studying the convergence of asynchronous fixed-point iterations involving maximum norm pseudo-contractions. We present a powerful approach for characterizing the rate of convergence of totally asynchronous iterations, where both the update intervals and communication delays may grow unbounded. When specialized to partially asynchronous iterations (where the update intervals and communication delays have a fixed upper bound), or to particular classes of unbounded delays and update intervals, our approach allows to quantify how the degree of asynchronism affects the convergence rate. In addition, we use our results to analyze the impact of asynchrony on the convergence rate of discrete-time power control algorithms in wireless networks.

The thesis finally proposes an asynchronous parallel algorithm that exploits multiple processors to solve regularized stochastic optimization problems with smooth loss functions. The algorithm allows the processors to work at different rates, perform computations independently of each other, and update global decision variables using out-of-date gradients. We characterize the iteration complexity and the convergence rate of the proposed algorithm, and show that these compare favourably with the state of the art. Furthermore, we demonstrate that the impact of asynchrony on the convergence rate of the algorithm is asymptotically negligible, and a near-linear speedup in the number of processors can be expected.

Populär sammanfattning

Tidsfördröjningar uppstår ofta i tekniska system: det tar tid för två ämnen att blandas, det tar tid för en vätska att rinna från ett kärl till ett annat, och det tar tid att överföra information mellan delsystem. Dessa tidsfördröjningar leder ofta till försämrad systemprestanda och ibland även till instabilitet. Det är därför viktigt att utveckla teori och ingenjörsmetodik som gör det möjligt att bedöma hur tidsfördröjningar påverkar dynamiska system.

I den här avhandlingen presenteras flera bidrag till detta forskningsområde. Fokus ligger på att karaktärisera hur tidsfördröjningar påverkar konvergenshastigheten hos olinjära dynamiska system. I kapitel 3 och 4 behandlar vi olinjära system vars tillstånd alltid är positiva. Vi visar att stabiliteten av dessa positiva system är oberoende av tidsfördröjningar och karaktäriserar hur konvergenshastigheten hos olinjära positiva system beror på tidsfördröjningarnas storlek. I kapitel 5 betraktar vi iterationer som är kontraktionsavbildningar, och analyserar hur deras konvergens påverkas av begränsade och obegränsade tidsfördröjningar. I avhandlingens sista kapitel föreslår vi en asynkron algoritm för stokastisk optimering vars asymptotiska konvergenshastighet är oberoende av tidsfördröjningar i beräkningar och i kommunikation mellan beräkningselement.

To Fereshteh, Kamran,
and Leila

Acknowledgements

A PhD thesis is a long journey that cannot be led to success without the help and support of many people.

First and foremost, I would like to extend my utmost gratitude to my main advisor, Mikael Johansson, for his endless support, care, friendship, and encouragement that helped me overcome the hurdles of research. I have learned a variety of research skills from Mikael including scientific writing, presentation skills, and (hopefully) good taste in choosing problems. I leave with a wonderful memory that every time I left his office after a meeting, I felt inspired. Thanks Mikael!

Second, I would like to thank my co-advisor, Alexandre Proutiere, for fruitful discussions on optimization and game theory. I also feel fortunate to have had the opportunity to closely collaborate with Ather Gattami, my former co-supervisor, at the beginning of my studies. I would like to thank him for motivating me to work on decentralized control problems.

During my time at KTH, I collaborated and wrote papers with great people. Thanks to Arda, Assad, Burak, Euhanna, Jie and Themis. I am specially grateful to Themis for the many interesting technical conversations with him, and to Arda for using his unbelievable programming skills. I also want to thank Arda, Bart, Sadegh, and Themis for reviewing parts of this thesis, and Niclas for helping me writing the Swedish summary.

I have many good memories from my times at the Automatic Control department. A big thanks to all my present and former colleagues for making a friendly working environment: António G. for all the interesting conversations we had on topics ranging from culture to politics, Arda for our relaxing lunches and dinners before deadlines, Burak for stimulating and fun talks, Demia for organizing challenging social events, Kaveh for all the awesome time in Frescatihallen, Meng for always being in a good mood, Themis for the days we were working together, and Afrooz, Antonio A., Assad, Bart, Christian, Dimitris, Emma, Euhanna, Farhad, Håkan, Hossein, Jie, José, Liqun, Mariette, Martin, Mohamed, Mohammadreza, Niclas, Niklas, Olle, Patricio, Riccardo, Sadegh, Sebastian, Sindri, Stefan, Valerio, Yuzhe, Zhenhua. I apologize in advance to anyone whom I forgot to mention.

I have learned a great deal from the professors at the Automatic Control department and I thank them all. Thanks also to Anneli, Gerd, Hanna, and Karin for your great spirit and help with all kinds of administrative issues.

I want to thank all my friends here in Sweden and back in Iran, especially Milad, Omid, and Shahab. Special thanks goes to Tina for her generous support when I moved to Sweden. I also want to thank Amir for all the joy and funny moments we shared for more than 20 years.

I would like to express my deepest gratitude to my parents, Fereshteh and Kamran, for their truly unconditional love and support. And finally Leila (*ea*), my ever patient, ever supportive companion, thank you for putting up with me and making me feel comfortable about doing all the things I have done.

Hamid Reza Feyzmahdavian
Stockholm, December 2015.

Notations

$:=$	Definition
$\mathbb{N}$	Set of all natural numbers
$\mathbb{N}_0$	Set of all natural numbers including zero
$\mathbb{R}$	Set of all real numbers
$\mathbb{R}_+$	Set of all nonnegative real numbers
$\mathbb{R}^n$	Set of all real vectors with n components
x_i	The i^{th} element of the vector $x \in \mathbb{R}^n$
$x \geq y$	$x_i \geq y_i$ for all i
$x > y$	$x_i > y_i$ for all i
$\mathbb{R}^n_+$	The set of all vectors in $\mathbb{R}^n$ with nonnegative entries $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x_i \geq 0,\ 1 \leq i \leq n\}$
$\langle x, y\rangle$	Inner product of two vectors x and y
$\mathbb{R}^{n\times n}$	Set of all real matrices of dimension $n \times n$
$A^\top$	Transpose of the matrix A
$\mathbf{1}$	Column vector with all elements equal to one
$\mathbf{0}$	Column vector with all elements equal to zero
I_n	Identity matrix in $\mathbb{R}^{n\times n}$
$\\|\cdot\\|_p$	The vector p-norm
$\\|\cdot\\|_*$	Dual norm to the norm $\\|\cdot\\|$, $\\|y\\|_* := \sup_{\\|x\\|\leq 1} \langle x, y\rangle$
$\nabla f(x)$	Gradient of f evaluated at x
$\mathbb{E}[x]$	Expected value of the random variable x
$\mathcal{C}\big([a,b], \mathbb{R}^n\big)$	Space of all continuous functions on $[a,b]$ taking values in $\mathbb{R}^n$
$D^+h(t)\big\rvert_{t=t_0}$	Upper-right Dini-derivative of a continuous function h at $t = t_0$

Vectors are written in lower case letters and matrices in capital letters.

Contents

Chapter 1

Introduction

Large-scale complex dynamical systems arise in a broad spectrum of applications such as biological and ecological systems, chemical processes, electrical power systems, communication networks, transportation systems, and urban water supply networks. These systems are highly interconnected and composed of large number of interacting subsystems that exchange material, energy, or information. In practice, propagation of physical quantities between subsystems may take place over large distances and is not instantaneous. Hence, communication delays are inevitably omnipresent in distributed systems. Even if communication delays are negligible, computational delays are strongly involved in complex systems. This is mainly due to that the subsystems can be heterogeneous (have non-identical dynamics) and require different computational times for the state evaluation. Therefore, in order to accurately describe and predict the behaviour of real-world large-scale systems, mathematical models of such systems must include time-delays.

Mathematical models of dynamical systems with time-delays, also called *time-delay systems*, take into account the dependence of the evolution of a system on the history of state variables. The dynamics of time-delay systems are much richer than their non-delayed counterparts. If a system without time-delays can be described by ordinary differential equations, the system with delays belong to the class of functional differential equations which are infinite-dimensional. The stability analysis of time-delay systems has been an active area of research in control engineering for more than 60 years. Existing results regarding this topic can be classified into two major categories: (i) *delay-independent* stability and (ii) *delay-dependent* stability. The delay-independent criteria guarantee stability regardless of the size of delays, whereas the delay-dependent criteria include information on the delay margin and provide a maximal allowable delay that can be tolerated by the system. Delay-dependent conditions are often less conservative, particulary, when the delay is small. On the other hand, delay-independent conditions are simpler and more appropriate to apply in the case that the delay is unknown, arbitrarily large, or unbounded.

Delay-dependent and delay-independent stability analysis of large-scale systems are very challenging especially when the subsystems have nonlinear dynamics and delays are time-varying. An effective approach to overcome these difficulties is to exploit specific structures for complex systems. There is a major on-going research effort in this direction, and this thesis is a part of that effort. In particular, the main objective of this thesis is to investigate delay-independent stability of a significant class of nonlinear systems, called positive systems, and study delay-dependent stability of asynchronous algorithms for stochastic optimization.

1.1 Time-delay Positive Systems

Positive systems are dynamical systems whose state variables are constrained to be nonnegative for all time whenever the initial conditions are nonnegative. Since the state variables of many real-world processes represent quantities that may not have meaning unless they are nonnegative, positive systems arise frequently in mathematical modelling of engineering problems [1]. Examples of nonnegative quantities are population levels of species in ecological systems [2], transmit power of mobile users in wireless networks [3], and concentration of substances in chemical processes [4]. Due to their importance and wide applicability, a large body of literature has been concerned with the analysis and control of positive systems (see, e.g., [5–13] and references therein).

In the following, two examples are used to illustrate the presence of time-delays in positive systems.

Example 1.1. Consider a wireless network where n mobile users communicate over the same frequency band. Since concurrent transmissions interfere with each other, users must transmit with sufficient power to overcome the interference caused by the others. Power control algorithms allow us to find transmit powers such that each user has a successful connection. In order to study this practical problem in wireless communication, power control algorithms are described by dynamical systems whose states are transmit power of users. For instance, continuous-time power control algorithms are described by

$$\dot{p}_i(t) = k_i\big(-p_i(t) + I_i(p(t))\big), \quad i = 1, \dots, n. \tag{1.1}$$

Here, $p_i(t)$ is the transmit power of user i at time t, $I_i : \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ is the interference function modeling the interference and noise experienced by the intended receiver of user i, and k_i is a positive constant [14]. Since the transmit power is a nonnegative quantity, the power control algorithm (1.1) defines a positive system.

In practice, there will always be a signaling delay associated with transmitting the perceived interference at the transmitter to the receiver, so that it can adjust the power according to the power control law. Consequently, a realistic analysis of continuous-time power control algorithms must consider heterogeneous time-varying delays. More precisely, the continuous-time power control algorithm (1.1), when the time-delays are introduced becomes

$$\dot{p}_i\big(t\big) = k_i\Big(-p_i(t) + I_i\big(p_1(t-\tau_1^i(t)), \dots, p_n(t-\tau_n^i(t))\big)\Big),$$

where $\tau_j^i(t)$ is the communication delay from user j to the intended receiver of user i at time t. The physical constraint that the transmit power should be nonnegative ($p_i(t) \geq 0$ for all $t \geq 0$) implies that asynchronous power control algorithms are also positive systems. ■

Example 1.2. A key challenge for health workers engaged in designing effective treatment strategies is to understand the underlying mechanisms of biological processes and epidemics. Considering epidemics and diseases as dynamical processes can reveal such mechanisms [15].

Time-delay positive systems are often used in mathematical modeling of hematology dynamics. For example, let x represent the circulating cell population of a certain type of blood cell, and let λ be the cell-loss rate in the circulation. The dynamics of the number of circulating cells in one compartment can be described by

$$\dot{x}(t) = -\lambda x(t) + G\big(x(t-\tau)\big),$$

where the function G denotes the flux of cells from the previous compartment, and the delay τ represents the average length of time required to go through the compartment. This time-delay system is positive since the circulating cell population is a nonnegative quantity. ■

For general dynamical systems, time-delays may render an otherwise stable system unstable. However, recent results have shown that if a positive linear system without delay is asymptotically stable, the corresponding system with either constant or bounded time-varying delays is also asymptotically stable. This means that the stability condition for a positive linear system with time-delays is the same as the stability condition for the delay-free system.

While many important positive systems such as power control algorithms and population dynamics are nonlinear, the theory for time-delay positive nonlinear systems is considerably less well-developed. In this thesis, we therefore investigate the following questions:

- Does the delay-independent property of positive linear systems hold also for positive nonlinear systems?
- Can we derive necessary and sufficient conditions for delay-independent stability of positive nonlinear systems which include previous results on positive linear systems as special cases?
- How do the maximum delay bound and the rate at which delays grow large affect the decay rate of positive systems?
- For what classes of unbounded time-varying delays is stability of positive linear systems insensitive to time-delays?

1.2 Asynchronous Algorithms for Stochastic Optimization

Asynchronous computation has a long history in optimization. Many early results were unified and significantly extended in the influential book by Bertsekas and Tsitsklis [16]. Renewed interest in the theoretical understanding and practical implementation of asynchronous optimization algorithms has been generated by recent advances in distributed and parallel computing technologies. In this thesis, we particularly focus on asynchronous algorithms for stochastic optimization.

The problem of stochastic optimization is the minimization of the expectation of a stochastic loss function:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\; f(x) := \mathbb{E}_\xi\big[F(x,\xi)\big] = \int_\Xi F(x,\xi)d\mathcal{P}(\xi). \tag{1.2}$$

Here, x is the decision vector, and ξ is a random variable whose probability distribution $\mathcal{P}$ is supported on a set $\Xi \subseteq \mathbb{R}^m$. A difficulty when solving stochastic optimization problems is that the distribution $\mathcal{P}$ is often unknown, so the expectation (1.2) cannot be computed. This situation occurs frequently in data-driven applications such as machine learning. One such application is *logistic regression* for classification tasks: we are given a set of observations

$$\Big\{\xi_j = \big(\xi_j^{(1)}, \xi_j^{(2)}\big) \mid \xi_j^{(1)} \in \mathbb{R}^n,\ \xi_j^{(2)} \in \{-1,+1\},\ j = 1,\dots,J\Big\},$$

drawn from an unknown distribution $\mathcal{P}$, and we want to learn a linear classifier to describe the relation between $\xi_j^{(1)}$ and $\xi_j^{(2)}$. To this end, we can solve the minimization problem (1.2) with

$$F(x,\xi) = \log\big(1 + \exp\big(-\xi^{(2)}\langle\xi^{(1)}, x\rangle\big)\big).$$

Stochastic gradient methods have become extremely popular for solving stochastic optimization problems [17–22]. Their popularity comes mainly from the fact that they are easy to implement and have low computational cost per iteration. With stochastic gradient methods, we do not assume knowledge of f (or of $\mathcal{P}$), but access to a stochastic oracle. Each time the oracle is queried with an $x \in \mathbb{R}^n$, it randomly selects ξ and returns $\nabla_x F(x,\xi)$, which is an unbiased estimate of $\nabla f(x)$. Classical stochastic gradient methods iteratively update the current vector $x(k)$ by computing $g(k) = \nabla_x F(x(k),\xi)$ and performing the update

$$x(k+1) = x(k) - \gamma(k)g(k), \quad k \in \mathbb{N}_0,$$

where $\gamma(k)$ is a positive step-size.

Stochastic gradient methods are inherently *serial* in the sense that gradient computations take place on a single processor which has access to the whole dataset and updates iterations sequentially, *i.e.*, one after another. In many emerging applications, such as large-scale machine learning and statistics, the size of dataset is so huge (in the Terabytes to Petabytes range) that it cannot fit on one computer.

For instance, a social network with 100 million users and 1KB data per user has 100GB [23]. The immense growth of available data has caused a strong interest in developing *parallel* optimization algorithms which are able to conveniently and efficiently split the data and distribute the computation across multiple processors or multiple computer clusters (see, *e.g.*, [24–31] and references therein). The performance of Google's DistBelief model [32] and Microsoft's Project Adam [33] have proven that parallel stochastic gradient methods are remarkably effective in real-world machine learning problems such as training deep learning systems. For example, while training a neural network for the ImageNet task with 16 million images may take about two weeks on a modern GPU, Google's DistBelief model can successfully utilize 16,000 cores in parallel and train the network for three days [32].

A common parallel implementation of stochastic gradient methods is the master-worker architecture in which several worker processors compute stochastic gradients in parallel based on their portions of the dataset while a master processor stores the decision vector and updates the current iterate. The workers communicate only with the master to retrieve the updated decision vector. The master-worker implementation can be executed in two ways: *synchronous* and *asynchronous*. In the synchronous case, the master will perform an update and broadcast the new decision vector to the workers when it collects stochastic gradients from *all* the workers (cf. Figure 1.1).

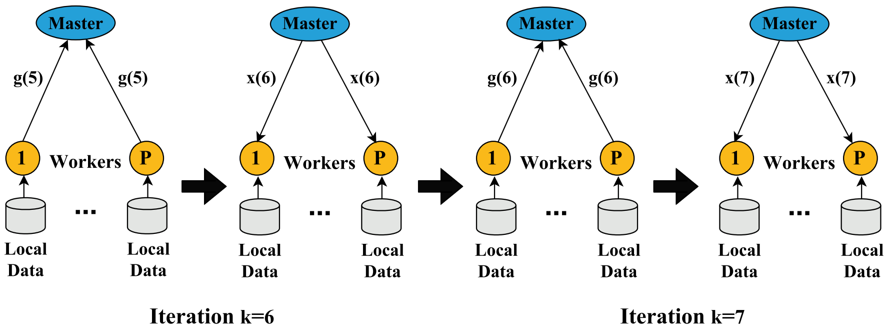


Figure 1.1: Synchronous implementation of a master-worker architecture with one master and P workers. At each iteration, the workers have to be synchronized with each other such that all the stochastic gradients $g(k) = \nabla F(x(k), \xi)$ are computed at the same vector $x(k)$. Furthermore, in order to update the decision vector, the master needs to wait until all the workers send their gradient computations.

Due to different computational capabilities, imperfect workload partition or interference by other running jobs, some workers may evaluate stochastic gradients slower than others. Since the master should wait for all the workers to finish their computations, synchronous parallel methods often suffer from the straggler problem [34], in which the algorithm can move forward only at the pace of the slowest worker. The need for global synchronization also make such methods fragile to many types of failures that are common in distributed computing environments. For example, if one processor fails throughout the execution of the algorithm or is disconnected from the network connecting the processors, the algorithm will come to an immediate halt. This becomes another bottleneck for synchronous parallel methods.

In contrast to synchronous parallel algorithms, asynchrony allows the workers to compute gradients at different rates without synchronization, and lets the master perform updates using out-of-date gradients. In other words, there is no need for workers to wait for each other to finish the gradient computations and the master can update the decision vector once it receives stochastic gradients even from one worker (cf. Figure 1.2). Some advantages that we can gain from asynchronous implementations of optimization algorithms:

1. Reduced idle time of processors;
2. More iterates executed by fast processors;
3. Alleviated congestion in inter-process communication;
4. Robustness to individual processor failures.

However, on the negative side, asynchrony runs the risk of rendering an otherwise convergent algorithm divergent. Asynchronous optimization algorithms often converge under more restrictive conditions than their synchronous counterparts. Thus, tuning an algorithm to withstand large amounts of asynchrony will typically result in unnecessarily slow convergence if the actual implementation is synchronous.

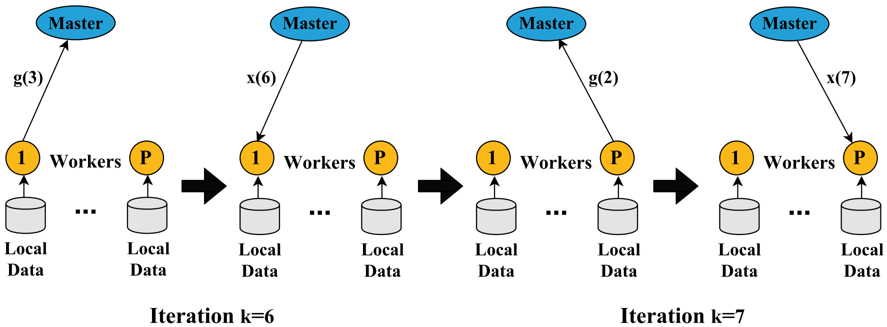


Figure 1.2: Asynchronous implementation of a master-worker architecture with one master and P workers. The workers evaluate stochastic gradient vectors independently of each other without the need for coordination or synchronization. When a small subset of the workers return their (possibly) out-of-date computations, the master can perform an update and pass the updated decision vector back to the workers.

In this thesis, we study asynchronous stochastic gradient methods for solving *regularized stochastic optimization* (also referred to as *stochastic composite optimization*) problems, which can be written in the form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \mathbb{E}_\xi\big[F(x,\xi)\big] + \Psi(x).$$

The role of the regularization term $\Psi(x)$, which may be non-differentiable, is to impose solutions with certain preferred structures. For example, $\Psi(x) = \lambda\|x\|_1$ with $\lambda > 0$ is often used to promote sparsity in solutions. Regularized stochastic optimization problems arise in many applications in machine learning, signal processing, and statistical estimation. Examples include Tikhonov and elastic net regularization, Lasso, sparse logistic regression, and support vector machines [35–37].

We focus on the following questions related to the asynchronous stochastic optimization:

- What is the update rule of an asynchronous parallel algorithm for regularized stochastic optimization? How should we tune the algorithm parameters so that the convergence is guaranteed in the face of asynchronism?
- What is the impact of asynchrony on the convergence rate of an asynchronous parallel algorithm for regularized stochastic optimization?
- Is it possible for an asynchronous parallel optimization algorithm to enjoy linear speedup in the number of processors?

1.3 Outline and Contributions

This section presents the outline and contributions of the thesis in detail. A more thorough description and the corresponding related work are provided in each chapter.

Chapter 2

In this chapter, we present mathematical background on which the rest of the thesis is built. In particular, we describe positive systems and introduce several classes of positive nonlinear systems. Then, we discuss some results concerning fixed point iterations and contraction mappings. We also review basic convexity notions and some first-order methods for solving smooth and nonsmooth convex optimization problems.

Chapter 3

In this chapter, we establish asymptotic stability and estimate the decay rate of a particular class of positive nonlinear systems which includes positive linear systems as a special case. More specifically, we present a set of necessary and sufficient conditions for delay-independent stability of continuous-time positive systems whose vector fields are cooperative and homogeneous. We show that global asymptotic stability of such positive monotone systems is independent of the magnitude and variation of time-delays. For various classes of bounded and unbounded time-varying delays, we derive explicit expressions that allow us to quantify the impact of delays on the decay rate. We demonstrate that the best decay rate of positive linear systems that our results provide can be found via convex optimization. Furthermore, we provide the corresponding counterparts for discrete-time positive nonlinear systems whose vector fields are order-preserving and homogeneous.

The chapter is based on the following publications:

- H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Asymptotic stability and decay rates of homogeneous positive systems with bounded and unbounded delays. *SIAM Journal on Control and Optimization*, 52(4):2623–2650, 2014.

- H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Exponential stability of homogeneous positive systems of degree one with time-varying delays. *IEEE Transactions on Automatic Control*, 59:1594–1599, 2014.

- H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Asymptotic stability and decay rates of positive linear systems with unbounded delays. *In Proceeding of IEEE Conference on Decision and Control Conference (CDC)*, pages 6337–6342, 2013.

- H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. On the rate of convergence of continuous-time linear positive systems with heterogeneous time-varying delays. *In Proceeding of European Control Conference (ECC)*, pages 3372–3377, 2013.

Chapter 4

The aim of this chapter is to study delay-independent stability of general positive monotone (not necessarily homogenous) systems with heterogeneous time-varying delays. We derive a set of necessary and sufficient conditions for asymptotic stability of positive monotone systems with bounded delays. Under the additional assumption of sub-homogeneity of vector fields, which includes homogeneous vector fields as a special case, we prove that a sub-homogeneous positive monotone system with time-varying delays is globally asymptotically stable if and only if the corresponding delay-free system is globally asymptotically stable. We also show how our results can be used to analyze the delay-independent stability of continuous-time power control algorithms in wireless networks.

The following publications provide the cornerstones for this chapter.

- H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Sub-homogeneous positive monotone systems are insensitive to heterogeneous time-varying delays. *In Proceeding of 21st International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, pages 317–324, 2014.

- H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Stability and performance of continuous-time power control in wireless networks. *IEEE Transactions on Automatic Control*, 59(8):2012–2023, 2014.

- H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Asymptotic and exponential stability of general classes of continuous-time power control laws in wireless networks. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 49–54, 2013.

Chapter 5

This chapter presents a unifying convergence result for asynchronous fixed point iterations involving pseudo-contractions in the block-maximum norm. Contrary to previous results in the literature which only established asymptotic convergence or investigated decay rates of simplified models of asynchronism, our results allow to characterize the convergence rates for various classes of update intervals and information delays. Furthermore, we use our main results to analyze the impact of asynchrony on the convergence rate of discrete-time power control algorithms in wireless networks.

The chapter is founded on the publications below.

- H. R. Feyzmahdavian and M. Johansson. On the convergence rates of asynchronous iterations. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 153–159, 2014.

- H. R. Feyzmahdavian, M. Johansson, and T. Charalambous. Contractive interference functions and rates of convergence of distributed power control laws. *IEEE Transactions on Wireless Communications*, 11(12):4494–4502, 2012.

- H. R. Feyzmahdavian, M. Johansson, and T. Charalambous. Contractive interference functions and rates of convergence of power control laws. *In Proceeding of IEEE International Conference on Communications (ICC)*, pages 5906–5910, 2012.

Chapter 6

In this chapter, we propose an asynchronous parallel algorithm for regularized stochastic optimization problems with smooth loss functions. We characterize the iteration complexity and the convergence rate of the proposed algorithm for convex and strongly convex regularization functions. We show that the asymptotic penalty in convergence rate of the algorithm due to asynchrony is asymptotically negligible and a near-linear speedup in the number of processors can be expected.

The following publications contribute to this chapter.

- H. R. Feyzmahdavian, A. Aytekin, and M. Johansson. An asynchronous mini-batch algorithm for regularized stochastic optimization. *Submitted to IEEE Transactions on Automatic Control*, 2015.

- H. R. Feyzmahdavian, A. Aytekin, and M. Johansson. An asynchronous mini-batch algorithm for regularized stochastic optimization. *To appear in IEEE Conference on Decision and Control (CDC)*, 2015.

Chapter 7

In this chapter, we summarize the thesis and discuss the results. We further outline possible directions to be taken in order to extend the work started with this thesis.

Other Contributions

For consistency of the thesis structure, the following publications by the author are not covered in the thesis.

- H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Delay-independent stability of cone-invariant monotone systems. *To appear in IEEE Conference on Decision and Control (CDC)*, 2015.

- B. Demirel, H. R. Feyzmahdavian, E. Ghadimi, and M. Johansson. Stability analysis of discrete-time linear systems with unbounded stochastic delays. *To appear in 5th IFAC Workshop on Distributed Estimation and Control of Networked Systems (NECSYS)*, 2015.

- E. Ghadimi, H. R. Feyzmahdavian, and M. Johansson. Global convergence of the heavy-ball method for convex optimization. *In Proceeding of European Control Conference (ECC)*, pages 310–315, 2015.

- J. Lu, H. R. Feyzmahdavian, and M. Johansson. A dual coordinate descent algorithm for multi-agent optimization. *In Proceeding of European Control Conference (ECC)*, pages 715–720, 2015.

- H. R. Feyzmahdavian, A. Aytekin, and M. Johansson. A delayed proximal gradient method with linear convergence rate. *In Proceeding of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2014.

- A. Aytekin, H. R. Feyzmahdavian, and M. Johansson. Asynchronous incremental block-coordinate descent. *In Proceeding of Annual Allerton Conference on Communication, Control, and Computing*, pages 19–24, 2014.

- H. R. Feyzmahdavian, A. Gattami, and M. Johansson. Distributed output-feedback LQG control with delayed information sharing. *In Proceeding of 3rd IFAC Workshop on Distributed Estimation and Control of Networked Systems (NECSYS)*, pages 192–197, 2012.

- H. R. Feyzmahdavian, A. Alam, and A. Gattami. Optimal distributed controller design with communication delays: Application to vehicle formations. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 2232–2237, 2012.

Chapter 2

Background

In this chapter, we briefly review the mathematical background of the thesis. The outline of the chapter is as follows. In Section 2.1, we describe positive systems and introduce useful definitions and results. We then discuss classes of cooperative, homogeneous, and sub-homogeneous systems in the context of positive systems. In Section 2.2, we present some theory regarding contraction mappings. Section 2.3 introduces important notions for convex optimization and reviews first-order methods relevant for the thesis.

2.1 Positive Systems

Consider the nonlinear autonomous system

$$\begin{cases} \dot{x}(t) = f\big(x(t)\big), & t \geq 0, \\ x(0) = x_0, \end{cases} \tag{2.1}$$

where $x(t) \in \mathbb{R}^n$ is the system state, $f : \mathcal{S} \to \mathbb{R}^n$ is continuously differentiable on $\mathcal{S} \subseteq \mathbb{R}^n$, and $x_0 \in \mathcal{S}$ is the initial condition. We denote the solution of (2.1) starting from x_0 by $x(t, x_0)$.

Definition 2.1. The dynamical system (2.1) is called positive if starting from any initial condition in the positive orthant, the trajectory of the system will remain in the positive orthant. That is,

$$x_0 \in \mathbb{R}^n_+ \Longrightarrow x(t, x_0) \in \mathbb{R}^n_+, \quad \forall t \geq 0.$$

This definition states that the positive orthant in $\mathbb{R}^n$ is an invariant set for positive systems. Positivity of nonlinear systems is readily verified using the following result.

Theorem 2.1 ([**58, Proposition 2.1**]). *Assume that $\mathbb{R}^n_+ \subseteq \mathcal{S}$. The dynamical system* (2.1) *is positive if and only if*

$$\forall x \in \mathbb{R}^n_+ \; : \; x_i = 0 \Longrightarrow f_i(x) \geq 0. \tag{2.2}$$

Intuitively, the positivity condition (2.2) means that at the boundary of the positive orthant $\mathbb{R}^n_+$, the vector field f is either zero or points toward the interior of $\mathbb{R}^n_+$, thus preventing the trajectories to leave $\mathbb{R}^n_+$.

Example 2.1. Consider the Lotka-Volterra equations

$$\begin{aligned}\dot{x}_1(t) &= \alpha x_1(t) - \beta x_1(t) x_2(t),\\ \dot{x}_2(t) &= \delta x_1(t) x_2(t) - \gamma x_2(t),\end{aligned}$$

which describe the population model for two species that interact in a predator-prey relationship [59]. Here, x_1 denotes the number of prey, x_2 denotes the number of predators, and α, β, γ, and δ are positive constants. In terms of (2.1),

$$f(x_1, x_2) = \begin{bmatrix}\alpha x_1 - \beta x_1 x_2\\ \delta x_1 x_2 - \gamma x_2\end{bmatrix}.$$

For any $(x_1, x_2) \in \mathbb{R}^2_+$, we have

$$\begin{aligned}x_1 = 0 &\Longrightarrow f_1(x_1, x_2) = 0 \geq 0,\\ x_2 = 0 &\Longrightarrow f_2(x_1, x_2) = 0 \geq 0.\end{aligned}$$

Therefore, according to Theorem 2.1, this system is positive. ■

The positive orthant plays an important role in the stability analysis of positive systems. While the vector field describing the evolution of a positive system may have equilibrium points outside the positive orthant, from the viewpoint of applications, it is only the stability of the nonnegative equilibria that are of interest. Therefore, it is natural to define the global asymptotic stability of a positive system by requiring that its equilibrium in $\mathbb{R}^n_+$ is asymptotically stable for any nonnegative initial condition $x_0 \in \mathbb{R}^n_+$, instead of for any $x_0 \in \mathbb{R}^n$. This means that an equilibrium which is not stable with respect to the whole $\mathbb{R}^n$ can be globally asymptotically stable with respect to the positive orthant. The following example illustrates this issue.

Example 2.2. Consider a scalar system described by (2.1) with

$$f(x) = -(x-1)(x+1)(x+3), \quad x \in \mathbb{R}. \tag{2.3}$$

Since $f(0) = 3 \geq 0$, it follows from Theorem 2.1 that (2.3) is positive. This system has three equilibrium points: $x^{\star(1)} = 1$, $x^{\star(2)} = -1$, and $x^{\star(3)} = -3$. It is easy to verify that $x^{\star(1)}$, which is the only equilibrium point in the positive orthant, is asymptotically stable for any initial condition $x_0 \in (-1, +\infty)$. As the trajectories of (2.3) starting from $x_0 \in (-\infty, -1)$ converges to $x^{\star(3)}$, $x^{\star(1)}$ is not globally asymptotically stable with respect to $\mathbb{R}$. However, for any nonnegative initial condition $x_0 \in \mathbb{R}_+$, $x(t, x_0)$ converges asymptotically to $x^{\star(1)}$. We conclude that $x^{\star(1)}$ is globally asymptotically stable with respect to $\mathbb{R}_+$. ■

2.1.1 Positive Linear Systems

In this subsection, we review some basic definitions and results concerning positive linear systems. Consider the linear time-invariant system

$$\begin{cases} \dot{x}(t) = Ax(t), \quad t \geq 0, \\ x(0) = x_0, \end{cases} \tag{2.4}$$

where $A \in \mathbb{R}^{n\times n}$, and $x_0 \in \mathbb{R}^n$. According to Theorem 2.1, positivity of the linear system (2.4) depends on the structure of A. The following definition introduces *Metzler* matrices.

Definition 2.2 (Metzler Matrix). A matrix $A \in \mathbb{R}^{n\times n}$ is called Metzler if its off-diagonal entries are nonnegative, *i.e.*, $a_{ij} \geq 0$ for all $i \neq j$, $i, j = 1, \ldots, n$.

Let $f(x) = Ax$, where $A \in \mathbb{R}^{n\times n}$ is Metzler. For each $i = 1, \ldots, n$,

$$f_i(x_1, \ldots, x_i = 0, \ldots, x_n) = \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij} x_j.$$

Since $a_{ij} \geq 0$ for all $i \neq j$, $f_i(x_1, \ldots, x_i = 0, \ldots, x_n) \geq 0$ for all $x \in \mathbb{R}^n_+$. This shows that if A is Metzler, then the positivity condition (2.2) is satisfied and, hence, the linear system (2.4) is positive. It is easy to verify that the requirement of being Metzler is also a necessary condition for positivity of linear systems.

Theorem 2.2 ([1, Theorem 2]). *The linear system* (2.4) *is positive if and only if A is Mezler.*

The linear system (2.4) has an equilibrium point at the origin. Stability properties of the origin can be characterized by the locations of the eigenvalues of the matrix A. It is well known that $x = \mathbf{0}$ is globally asymptotically stable if and only if all eigenvalues of A have negative real parts [60, Theorem 4.5]. The Lyapunov stability theorem provides an alternative condition to determine whether or not (2.4) is asymptotically stable. More precisely, (2.4) is globally asymptotically stable if and only if there exists a positive-definite matrix $P \in \mathbb{R}^{n\times n}$ such that

$$A^\top P + PA \text{ is negative definite.} \tag{2.5}$$

Such a matrix P corresponds to the quadratic Lyapunov function $V(x) = x^\top Px$, which is decreasing along trajectories of (2.4) [60, Theorem 4.6]. We can find numerically a positive-definite matrix P satisfying (2.5) by solving a convex optimization problem with $n(n+1)/2$ decision variables [61]. As for general linear systems, the existence of a quadratic Lyapunov function is necessary and sufficient for stability of positive linear systems. However, the next result shows that this type of Lyapunov functions has a simpler structure in the case of positive systems.

Theorem 2.3 (**[62, Proposition 1]**)**.** *Assume that $A \in \mathbb{R}^{n\times n}$ is Metzler. Then, for the positive linear system* (2.4)*, the following statements are equivalent:*

1. *The origin is globally asymptotically stable.*
2. *There exists a positive definite diagonal matrix P such that* (2.5) *holds.*
3. *There exists $w \in \mathbb{R}^n$ such that*

$$\begin{cases} w^\top A < \mathbf{0}, \\ w > \mathbf{0}. \end{cases} \tag{2.6}$$

4. *There exists $v \in \mathbb{R}^n$ such that*

$$\begin{cases} Av < \mathbf{0}, \\ v > \mathbf{0}. \end{cases} \tag{2.7}$$

This theorem suggests that to find a quadratic Lyapunov function for positive linear systems, it suffices to search for a diagonal matrix P satisfying (2.5). In this case, the asymptotic stability can be verified by a convex optimization problem involving only n decision variables. Theorem 2.3 also demonstrates that positive linear systems admit other classes of Lyapunov functions leading to necessary and sufficient conditions. Specifically, consider the linear Lyapunov function candidate $V(x) = w^\top x$, where w satisfies (2.6). It is clear that $V(\mathbf{0}) = 0$ and $V(x) > 0$ for all $x \in \mathbb{R}^n_+ - \{\mathbf{0}\}$. The derivative of V along the trajectories of (2.4) is given by

$$\dot{V}(x) = w^\top \dot{x} = w^\top Ax < 0, \quad \forall x \in \mathbb{R}^n_+ - \{\mathbf{0}\},$$

which implies that the origin is asymptotically stable. Similarly, if we can demonstrate the existence of a vector v satisfying (2.7), then

$$V(x) = \max_{1\le i\le n} \frac{x_i}{v_i},$$

is a Lyapunov function for the the positive linear system (2.4). Note that the necessary and sufficient stability conditions (2.6) and (2.7) are linear programming problems in w and v, respectively. In fact, the stability of (2.4) can be checked by finding a feasible solution to $2n$ linear inequalities over n variables.

Remark 2.1. The powerful properties of Metzler matrices presented in Theorem 2.3 can simplify stability analysis and control design problems for positive linear systems [63–67]. For example, the design of structured static state-feedback controllers is known to be NP-Hard in general [68]. For positive linear systems, however, it was shown in [63] that finding structured $\mathcal{H}_\infty$ static state-feedback controllers can be reformulated as a semi-definite programming problem by employing diagonal Lyapunov functions. In [64], the synthesis of distributed output-feedback controllers for positive linear systems was solved in terms of linear programming. [65–67] provided necessary and sufficient tractable conditions for robust stability of uncertain positive linear systems in the l_1, l_2 and l_∞ gain setting, respectively.

2.1.2 Cooperative Positive Systems

Cooperative positive systems are a particular class of positive nonlinear systems which include positive linear systems as a special case. We first define cooperative vector fields.

Definition 2.3 (**Cooperativity**). A vector field $f : \mathcal{S} \to \mathbb{R}^n$ which is continuously differentiable on the convex set $\mathcal{S} \subseteq \mathbb{R}^n$ is said to be cooperative if the Jacobian matrix $\frac{\partial f}{\partial x}(a)$ is Metzler for all $a \in \mathcal{S}$. The dynamical system (2.1) is called cooperative if f is cooperative.

Loosely speaking, cooperativity means that an increase in the value of one component of the state variable causes an increase of the growth rates of all the other components. Cooperative systems occur in many biological models. The biological interpretation is that an increase of species i tends to increase the population growth rate of every other species j.

Example 2.3. Consider the system (2.1) with

$$f(x_1, x_2) = \begin{bmatrix} -x_1^2 + x_1 x_2^2 \\ 5x_1 - x_2^3 \end{bmatrix}.$$

The Jacobian matrix $\partial f / \partial x$ at a point (x_1, x_2) is given by

$$\frac{\partial f}{\partial x}(x_1, x_2) = \begin{bmatrix} -2x_1 + x_2^2 & 2x_1 x_2 \\ 5 & -3x_2^2 \end{bmatrix}.$$

Since the off-diagonal entries are nonnegative for all $(x_1, x_2) \in \mathbb{R}_+^2$, $\partial f / \partial x$ is Metzler over $\mathbb{R}_+^2$. Therefore, f is cooperative on $\mathbb{R}_+^2$. ∎

One important property of cooperative systems is that they are monotone [69, §3]. Monotone systems are those for which trajectories preserve a partial ordering on initial states. The formal definition of monotone systems is as follows.

Definition 2.4. The dynamical system (2.1) is called monotone in $\mathcal{S}$ if for any initial conditions $x_0, y_0 \in \mathcal{S}$, we have

$$x_0 \le y_0 \Longrightarrow x(t, x_0) \le x(t, y_0), \quad \forall t \ge 0.$$

The next result demonstrates that for continuously differentiable vector fields, monotone systems are necessarily cooperative.

Theorem 2.4 (**[70, Lemma 2.1]**). *For the dynamical system* (2.1)*, assume that f is continuously differentiable. Then,* (2.1) *is monotone if and only if f is cooperative.*

According to Theorem 2.4, the linear system (2.4) is monotone if and only if A Metzler. Hence, (2.4) is monotone if and only if it is positive.

Remark 2.2. The theory of monotone systems has been developed by Hirsch [71–73] and Smith [69]. In [74], the notion of monotone systems was extended to systems with inputs and outputs. Motivated by potential applications to a wide variety of areas such as molecular biology and chemical reaction networks, monotone systems have attracted considerable attention from the control community (see, e.g., [75–78]).

2.1.3 Homogeneous Positive Systems

In Chapter 3, we will deal with cooperative positive systems whose vector fields are homogeneous in the sense of the following definition.

Definition 2.5 (Homogeneity). Given an n-tuple $r = (r_1, \dots, r_n)$ of positive real numbers and $\lambda > 0$, the *dilation map* $\delta_\lambda^r(x) : \mathbb{R}^n \to \mathbb{R}^n$ is given by

$$\delta_\lambda^r(x) = \big(\lambda^{r_1} x_1, \dots, \lambda^{r_n} x_n\big).$$

When $r = (1, \dots, 1)$, the dilation map is called the *standard dilation map.* A vector field $f : \mathbb{R}^n \to \mathbb{R}^n$ is said to be homogeneous of degree p with respect to the dilation map $\delta_\lambda^r(x)$ if for all $x \in \mathbb{R}^n$ and all $\lambda > 0$,

$$f\big(\delta_\lambda^r(x)\big) = \lambda^p \delta_\lambda^r\big(f(x)\big).$$

Note that the linear system (2.4) is homogeneous of degree zero with respect to the standard dilation map since $f(\lambda x) = A(\lambda x) = \lambda A x$ for all $\lambda > 0$.

Example 2.4. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined as

$$f(x_1, x_2) = \begin{bmatrix} x_1^2 - 6x_1x_2^3 \\ 3x_1x_2 - x_2^4 \end{bmatrix}.$$

We show that f is homogeneous of degree $p = 3$ with respect to the dilation map $\delta_\lambda^r(x)$ with $r = (3, 1)$. We have

$$f\big(\delta_\lambda^r(x)\big) = f\big(\lambda^3 x_1, \lambda x_2\big) = \lambda^3 \begin{bmatrix} \lambda^3(x_1^2 - 6x_1x_2^3) \\ \lambda(3x_1x_2 - x_2^4) \end{bmatrix} = \lambda^3 \delta_\lambda^r\big(f(x)\big).$$

Therefore, $f\big(\delta_\lambda^r(x)\big) = \lambda^3 \delta_\lambda^r\big(f(x)\big)$ for all $x \in \mathbb{R}^2$ and all $\lambda > 0$. ∎

Returning to Theorem 2.3, we see that the positive linear system (2.4) is globally asymptotically stable if and only if there is some $v > \mathbf{0}$ satisfying $Av < \mathbf{0}$. The next result extends this stability property to positive nonlinear systems whose vector fields are cooperative, homogeneous and irreducible.

Theorem 2.5 ([79]). *Suppose that f is cooperative on $\mathbb{R}_+^n$ and homogeneous of degree $p \in \mathbb{R}_+$ with respect to the dilation map $\delta_\lambda^r(x)$. Suppose also that the Jacobian matrix $\partial f / \partial x(a)$ is irreducible for all $a \in \mathbb{R}_+^n - \{\mathbf{0}\}$. Then, the positive system* (2.1) *is globally asymptotically stable if and only if there exists $v > \mathbf{0}$ such that $f(v) < \mathbf{0}$.*

If f is homogeneous of degree zero with respect to the standard dilation map, the result of Theorem 2.5 still holds without requiring irreducibility assumption [80].

2.1.4 Sub-homogeneous Positive Systems

Another class of positive nonlinear systems that we will focus on in Chapter 4 is sub-homogeneous cooperative systems. The next definition introduces the concept of a sub-homogeneous vector field.

Definition 2.6 (**Sub-homogeneity**). A vector field $f : \mathbb{R}^n \to \mathbb{R}^n$ is called sub-homogeneous of degree $p \in \mathbb{R}_+$ with respect to the dilation map $\delta_\lambda^r(x)$ if for all $x \in \mathbb{R}^n$ and all $\lambda \geq 1$,

$$f\big(\delta_\lambda^r(x)\big) \leq \lambda^p \delta_\lambda^r\big(f(x)\big).$$

Example 2.5. Consider the following system

$$\dot{x}\big(t\big) = f\big(x(t)\big) + b, \quad t \geq 0,$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is homogeneous of degree $p \in \mathbb{R}_+$ with respect to the dilation map $\delta_\lambda^r(x)$, and $b \in \mathbb{R}_+^n$ is a constant control which allows to shift the equilibrium point from the origin to a point in the positive orthant [81]. Let $\hat{f}(x) = f(x) + b$. It follows from homogeneity of f that

$$\hat{f}\big(\delta_\lambda^r(x)\big) = f\big(\delta_\lambda^r(x)\big) + b = \lambda^p \delta_\lambda^r\big(f(x)\big) + b = \lambda^p \delta_\lambda^r\big(\hat{f}(x)\big) + b - \lambda^p \delta_\lambda^r(b).$$

Since $b_i \geq 0$ for each $i = 1, \ldots, n$, we have

$$b - \lambda^p \delta_\lambda^r(b) = \big((1 - \lambda^{p+r_1})b_1, \ldots, (1 - \lambda^{p+r_n})b_n\big) \leq \mathbf{0},$$

for all $\lambda \geq 1$. Therefore, $\hat{f}\big(\delta_\lambda^r(x)\big) \leq \lambda^p \delta_\lambda^r\big(\hat{f}(x)\big)$, which means that $\hat{f}$ is sub-homogeneous of degree p with respect to the dilation map $\delta_\lambda^r(x)$. ■

It is clear that every homogeneous vector field is also sub-homogeneous. However, the following simple example shows that the converse is, in general, not true.

Example 2.6. Consider $f(x) = x + 1$, $x \in \mathbb{R}$. Clearly, f is not homogeneous. However, for any $\lambda \geq 1$,

$$f(\lambda x) = \lambda x + 1 = \lambda(x + 1) + (1 - \lambda) \leq \lambda(x + 1) = \lambda f(x).$$

which implies that f is sub-homogeneous of degree zero. ■

2.1.5 Discrete-time Positive Systems

In the remainder of this section, we review some properties of discrete-time positive systems of the form

$$\begin{cases} x\big(k+1\big) & = f\big(x(k)\big), \quad k \in \mathbb{N}_0, \\ x(0) & = x_0. \end{cases} \tag{2.8}$$

Here, $x(k) \in \mathbb{R}^n$ is the state variable, $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuous on $\mathbb{R}^n$, and $x_0 \in \mathbb{R}^n$ represents the initial condition.

Definition 2.7. The discrete-time system (2.8) is said to be positive if for every nonnegative initial condition $x_0 \in \mathbb{R}^n_+$, the corresponding solution is nonnegative, *i.e.*, $x(k, x_0) \in \mathbb{R}^n_+$ for all $k \in \mathbb{N}$.

The following theorem provides a necessary and sufficient condition for positivity of discrete-time systems.

Theorem 2.6 (**[58, Proposition 2.11]**)**.** *The dynamical system* (2.8) *is positive if and only if* $f(x) \in \mathbb{R}^n_+$ *for all* $x \in \mathbb{R}^n_+$.

Consider now the discrete-time linear system

$$x(k+1) = Ax(k), \quad k \in \mathbb{N}_0, \tag{2.9}$$

where $A \in \mathbb{R}^{n \times n}$.

Definition 2.8 (**Nonnegative Matrix**)**.** A matrix $A \in \mathbb{R}^{n \times n}$ is called nonnegative if all of its elements are nonnegative, *i.e.*, $a_{ij} \geq 0$ for all $i, j = 1, \ldots, n$.

According to Theorem 2.6, the linear system (2.8) is positive if and only if $Ax \in \mathbb{R}^n_+$ for all $x \in \mathbb{R}^n_+$. This condition holds if and only if A is nonnegative. To see this, suppose one of the elements of A, a_{ij}, were negative. Then, for the nonnegative vector $x = (0, \ldots, 0, 1, 0, \ldots, 0)$ with the one in the i^{th} component, the j^{th} component of Ax would be a_{ij}, which is negative. It is also easy to verify the converse. Therefore, the linear system (2.8) is positive if and only if A is nonnegative.

Nonnegative matrices, which play a significant role in mathematical economics and Markov processes, have a remarkably rich theory. This theory has roots in the Perron-Frobenius theorem, which states that the spectral radius of a nonnegative matrix whose elements are strictly positive is an eigenvalue corresponding to the eigenvector with strictly positive components [82]. We summarize some well-known properties of nonnegative matrices. These conditions are useful when analyzing the stability and control of discrete-time positive linear systems.

Theorem 2.7 (**[62, Proposition 2]**)**.** *Assume that* $A \in \mathbb{R}^{n \times n}$ *is nonnegative. Then, for the positive linear system* (2.9)*, the following statements are equivalent:*

1. *The origin is globally asymptotically stable.*

2. *There exists a positive definite diagonal matrix* P *such that* $A^\top PA - P$ *is negative definite.*

3. *There exists* $w \in \mathbb{R}^n$ *such that*

$$\begin{cases} w^\top A < w, \\ w > \mathbf{0}. \end{cases}$$

4. There exists $v \in \mathbb{R}^n$ such that

$$\begin{cases} Av < v, \\ \quad v > \mathbf{0}. \end{cases}$$

This theorem shows that stable discrete-time positive linear systems admit three types of Lyapunov functions: the diagonal quadratic function $V(x) = x^\top Px$, the linear function $V(x) = w^\top x$, and the weighted infinity norm $V(x) = \max_{1 \le i \le n} x_i/v_i$.

Among different classes of discrete-time positive nonlinear systems, we will mainly deal with *order-preserving* systems.

Definition 2.9 (**Order-preserving System**). A vector field $f : \mathbb{R}^n \to \mathbb{R}^n$ is called order-preserving on $\mathbb{R}^n_+$ if $f(x) \le f(y)$ for any $x, y \in \mathbb{R}^n_+$ such that $x \le y$. The dynamical system (2.8) is said to be order-preserving if f is order-preserving.

Order-preserving systems are monotone in the sense that solutions starting at ordered initial conditions preserve the same ordering during the time evolution. More precisely, if $x_0 \le y_0$, then $x(k, x_0) \le x(k, y_0)$ for all $k \in \mathbb{N}_0$.

2.2 Contraction Mappings

Several iterative algorithms generate sequences $\{x(k)\}$ according to

$$x(k+1) = T(x(k)), \quad k \in \mathbb{N}_0, \tag{2.10}$$

where T is a mapping from $\mathcal{S} \subseteq \mathbb{R}^n$ into $\mathbb{R}^n$. If $\{x(k)\}$ converges to some $x^\star \in \mathcal{S}$ and T is continuous at $x^\star$, then

$$x^\star = T(x^\star). \tag{2.11}$$

Any vector $x^\star \in \mathcal{S}$ satisfying (2.11) is called a *fixed point* of T. Thus, a convergent iteration of the form (2.10) can be viewed as an algorithm for solving the fixed point problem $x = T(x)$. A classical optimization example is the iteration

$$x(k+1) = x(k) - \gamma \nabla f(x(k)),$$

where γ is a positive step-size, and $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. This iteration aims to solve the equation $x = x - \gamma \nabla f(x)$, or, equivalently, $\nabla f(x) = 0$, which is the optimality condition for the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n} \quad f(x).$$

We are typically interested in conditions that guarantee the convergence of the iteration (2.10) to some desirable fixed points. We are also interested in estimating the convergence rate of the sequence $\{x(k)\}$. A common approach for establishing the convergence of (2.10) is to verify that T is a contraction mapping.

Definition 2.10 (Contraction Mapping). A mapping T is called a *contraction mapping* with *contraction modulus* c if there is $c \in (0,1)$ such that

$$\|T(x') - T(x)\| \le c\,\|x' - x\|, \quad \forall x, x' \in \mathcal{S},$$

where $\|\cdot\|$ is some norm on $\mathcal{S}$.

Contraction mappings are automatically continuous. Furthermore, an attractive feature of such mappings is that they always have a unique fixed point and the corresponding iteration (2.10) converges to it at a linear rate.

Theorem 2.8 ([16, Proposition 3.1.1]). *Suppose that $T : \mathcal{S} \to \mathcal{S}$ is a contraction mapping with contraction modulus c and that $\mathcal{S}$ is a closed subset of $\mathbb{R}^n$. Then,*

- *T has a unique fixed point $x^\star \in \mathcal{S}$.*
- *For every initial condition $x(0) \in \mathcal{S}$, the sequence $\{x(k)\}$ generated by* (2.10) *converges linearly to $x^\star$. In particular,*

$$\|x(k) - x^\star\| \le c^k \|x(0) - x^\star\|, \quad k \in \mathbb{N}_0.$$

Note that if T is contractive but $\mathcal{S}$ is not closed, then T may not have a fixed point, see [16, Execersice 3.1.1]. In addition, the contractivity property is norm dependent, so that a mapping T might be a contraction for one norm but not for a different choice of norm. The following example illustrates this point.

Example 2.7. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear mapping defined as $T(x) = Ax$ with

$$A = \begin{bmatrix} 0.5 & 0.7 \\ 0 & -0.5 \end{bmatrix}.$$

Clearly, T is a contraction if $\|A\| < 1$ for some norm on $\mathbb{R}^2$. It is easy to verify that $\|A\|_1 = \|A\|_\infty = 1.2$ and $\|A\|_2 = 0.96$. Therefore, T is contractive with respect to the norm $\|\cdot\|_2$ while fails to be a contraction under $\|\cdot\|_1$ and $\|\cdot\|_\infty$ norms. ■

2.3 First-order Methods in Convex Optimization

Given an objective function $f : \mathbb{R}^n \to \mathbb{R}$ and a feasible set $\mathcal{S} \subseteq \mathbb{R}^n$, the goal of optimization is to find a point in $\mathcal{S}$ where f attains its minimum. Mathematically, an optimization problem can be formulated as

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{suject to} & x \in \mathcal{S}. \end{array} \tag{2.12}$$

When f is a convex function and $\mathcal{S}$ is a convex set, (2.12) is called a *convex optimization problem*. An important implication of convexity of the problem (2.12) is that any local minimum for (2.12) is also a global minimum [83, Proposition 2.1.2]. While the convexity assumption is restrictive, a vast number of design problems in engineering can be posed as, or at least approximated by, convex optimization problems [84]. This section provides an overview of useful definitions and describes some first-order methods relating to convex optimization. We refer the interested readers to [83–85] for a more thorough exposition.

2.3.1 Basic Definitions

We start with the definition of convex sets.

Definition 2.11 (**Convex Set**). A set $\mathcal{S} \subseteq \mathbb{R}^n$ is convex if

$$\theta x + (1-\theta)x' \in \mathcal{S},$$

for any $x, x' \in \mathcal{S}$ and for any $\theta \in [0,1]$.

The definition implies that the segment between any two points in a convex set must lie within the set (cf. Figure 2.1).

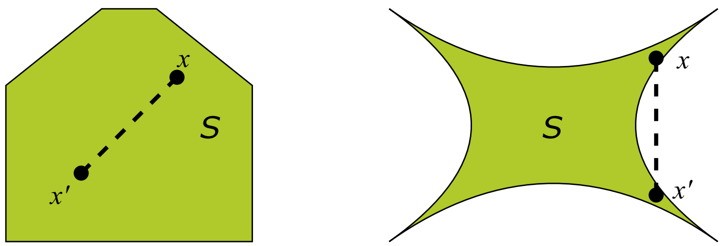


Figure 2.1: Illustration of the definition of a convex set. The set on the left is convex, but the set on the right is not.

Next, we define convex functions.

Definition 2.12 (**Convex Function**). A function $f : \mathcal{S} \to \mathbb{R}$ is convex if the domain $\mathcal{S}$ is convex and

$$f(\theta x + (1-\theta)x') \leq \theta f(x) + (1-\theta) f(x'),$$

for all $x, x' \in \mathcal{S}$ and for all $\theta \in [0,1]$.

Geometrically, convexity of f means that the graph of f between any two points in the domain lies below the line segment joining the points (cf. Figure 2.2).

Given a convex function f, if $f(x) < +\infty$ for at least one $x \in \mathcal{S}$ and $f(x) > -\infty$ for all $x \in \mathcal{S}$, then f is called *proper*. For example, the indicator function of a non-empty

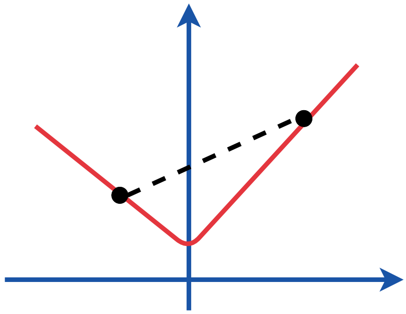


Figure 2.2: Example of a convex function.

closed convex set $C \subseteq \mathbb{R}^n$, given by

$$I_C(x) := \begin{cases} 0, & \text{if } x \in C, \\ +\infty, & \text{otherwise,} \end{cases}$$

is proper on $\mathbb{R}^n$. The following definition introduces *subgradients* of proper convex functions.

Definition 2.13 (**Subgradient**). For a proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, a vector $g \in \mathbb{R}^n$ is a subgradient of f at $x \in \mathbb{R}^n$ if

$$f(x') \geq f(x) + \langle g, x' - x \rangle, \quad \forall x' \in \mathbb{R}^n. \tag{2.13}$$

The set of all subgradients of f at x is called the subdifferential of f at x, and is denoted by $\partial f(x)$:

$$\partial f(x) := \left\{ g \in \mathbb{R}^n \mid f(x') \geq f(x) + \langle g, x' - x \rangle \text{ for all } x' \in \mathbb{R}^n \right\}.$$

The inequality (2.13) states that the subgradient gives an affine global underestimator of f. Note that if f is differentiable at x, then $\partial f(x)$ contains only one element [83, Proposition 4.2.2], namely $\partial f(x) = \{\nabla f(x)\}$.

A particular class of convex functions satisfies extra property, called *strong convexity*, that can significantly speedup the convergence of first-order methods.

Definition 2.14 (**Strongly Convex Function**). A function $f : \mathcal{S} \to \mathbb{R}$ over the convex set $\mathcal{S}$ is called μ-strongly convex with respect to the norm $\|\cdot\|$ if there exists a constant $\mu > 0$ such that

$$f(\theta x + (1-\theta)x') \leq \theta f(x) + (1-\theta) f(x') - \frac{\mu}{2}\theta(1-\theta)\|x' - x\|^2,$$

for all $x, x' \in \mathcal{S}$ and for all $\theta \in [0, 1]$.

The constant μ is called the *convexity parameter*. Clearly, every convex function is 0-strongly convex. Strong convexity gives us a quadratic lower bound to f at any point in the domain. More precisely, for all $x, x' \in \mathcal{S}$, we have

$$f(x') \geq f(x) + \langle g, x' - x \rangle + \frac{\mu}{2}\|x' - x\|^2, \quad \forall g \in \partial f(x).$$

2.3.2 First-order Methods

Numerical methods for solving optimization problems are commonly classified as *zero-*, *first-*, and *second-order methods* according to the derivative information which they require to compute. The zero-order methods (also referred to as *derivative-free methods*) use only the objective function values to direct the search for the optimum while the first-order methods require additional gradients (sub-gradients) of the objective function. Typical examples of zero- and first-order methods are the simplex method [86] and the gradient descent method [85], respectively. In second-order methods, the second derivative of the objective function, called the *Hessian*, is also used to construct search directions. Newton's method is a classical second-order method [85]. In this section, we are interested in first-order methods that solve the convex optimization problem (2.12).

One of the simplest first-order methods is the *subgradient projection method* which is among the earliest algorithms developed to solve (2.12). This method starts with some initial feasible vector $x(0) \in \mathcal{S}$, and updates the current iterate by taking a step along the negative subgradient direction and then, by projecting on the feasible set $\mathcal{S}$. Formally, the subgradient projection method proceeds according to

$$x\big(k+1\big) = \Pi_{\mathcal{S}}\big(x(k) - \gamma g(k)\big), \quad k \in \mathbb{N}_0, \tag{2.14}$$

where $x(k)$ is the current iterate, $g(k)$ is any subgradient of f at $x(k)$, γ is a positive step-size, and $\Pi_{\mathcal{S}}(\cdot)$ is the Euclidean projection operator onto the set $\mathcal{S}$, *i.e.*,

$$\Pi_{\mathcal{S}}(x) = \operatorname*{argmin}_{x' \in \mathcal{S}} \|x' - x\|_2^2.$$

Using the definition of $\Pi_{\mathcal{S}}(\cdot)$, one can verify that (2.14) can be rewritten as

$$x(k+1) = \operatorname*{argmin}_{x \in \mathcal{S}} \left\{ f(x(k)) + \big\langle g(k), x - x(k) \big\rangle + \frac{1}{2\gamma}\big\|x - x(k)\big\|_2^2 \right\}. \tag{2.15}$$

If $\mathcal{S}$ is closed and convex, the minimization problem (2.15) admits a unique solution [85, Theoorem 2.2.6]. We can interpret (2.15) as follows: next iterate is a unique minimizer of the linear approximation of f at the current iterate $x(k)$ plus a quadratic regularization term which penalizes deviations from $x(k)$. For differentiable objective functions, we can use the *gradient projection method* to solve (2.12), which has the form

$$x\big(k+1\big) = \Pi_{\mathcal{S}}\big(x(k) - \gamma \nabla f(x(k))\big), \quad k \in \mathbb{N}_0.$$

This method can be viewed as the specialization of the subgradient method applied to differentiable objective functions.

The subgradient and gradient projection methods can be easily extended to a non-Euclidean setting, by replacing the Euclidean squared distance in (2.15) with a general *Bregman distance function*. This was the idea behind the *mirror descent method* originated in [87] and developed further in [18, 88, 89]. We first define a Bregman distance function, also referred to as a *prox-function*.

Definition 2.15 (**Bregman Distance Function**). A function $\omega : \mathcal{S} \to \mathbb{R}$ is called a distance generating function with modulus $\mu_\omega > 0$ if ω is continuously differentiable and μ_ω-strongly convex with respect to $\|\cdot\|$ over the set $\mathcal{S} \subseteq \mathbb{R}^n$. Every distance generating function introduces a corresponding Bregman distance function given by

$$D_\omega(x, x') := \omega(x') - \omega(x) - \langle \nabla\omega(x), x' - x\rangle.$$

For example, choosing $\omega(x) = \frac{1}{2}\|x\|_2^2$, which is 1-strongly convex with respect to the l_2-norm over any convex set $\mathcal{S}$, would result in $D_\omega(x, x') = \frac{1}{2}\|x' - x\|_2^2$. Another common example of distance generating functions is the entropy function

$$\omega(x) = \sum_{i=1}^{n} x_i \log x_i,$$

which is 1-strongly convex with respect to the l_1-norm over the standard simplex

$$\Delta := \left\{ x \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} x_i = 1,\ x \geq 0 \right\},$$

and its associated Bregman distance function is

$$D_\omega(x, x') = \sum_{i=1}^{n} x'_i \log \frac{x'_i}{x_i}.$$

Remark 2.3. The strong convexity of the distance generating function ω always ensures that

$$D_\omega(x, x') \geq \frac{\mu_\omega}{2}\|x' - x\|^2, \quad \forall x, x' \in \mathcal{S},$$

and $D_\omega(x, x') = 0$ if and only if $x = x'$.

Remark 2.4. Throughout the thesis, there is no loss of generality to assume that $\mu_\omega = 1$. Indeed, if $\mu_\omega \neq 1$, we can choose the scaled function $\overline{\omega}(x) = \frac{1}{\mu_\omega}\omega(x)$, which has modulus $\overline{\mu}_\omega = 1$, to generate the Bregman distance function.

The update equation of the mirror descent method is

$$x(k+1) = \operatorname*{argmin}_{x \in \mathcal{S}} \left\{ f(x(k)) + \big\langle g(k), x - x(k) \big\rangle + \frac{1}{\gamma} D_\omega\big(x(k), x\big) \right\}, \quad k \in \mathbb{N}_0. \quad (2.16)$$

The subgradient method (2.14) is a special case of (2.16), when $\omega(x) = \frac{1}{2}\|x\|_2^2$ and, hence, $D_\omega(x, x') = \frac{1}{2}\|x' - x\|_2^2$. The main motivation to use a generalized distance generating function, instead of the usual Euclidean distance function, is to design optimization algorithms that can take advantage of the geometry of the feasible set (see, e.g., [89–91]). For example, when $\mathcal{S}$ is the standard simplex of large dimension n and f is convex and Lipschitz continuous on $\mathcal{S}$, the mirror descent method (2.16) with the entropy distance function can outperform the subgradient method (2.14) by a factor of $\mathcal{O}(\sqrt{n/\ln n})$ [18].

2.3.3 Iteration Complexity of Mirror Descent Method

We now discuss the convergence properties of the mirror descent method under two different assumptions.

Objective functions with bounded subgradients:
Suppose that there exists a constant $G \in (0, \infty)$ such that

$$\|g\|_* \le G, \quad \forall g \in \partial f(x),\ \forall x \in \mathcal{S}, \tag{2.17}$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. This constraint on the norm of subgradients implies that f is Lipschitz continuous over $\mathcal{S}$ with constant G [92, Lemma 14.7]. That is,

$$|f(x') - f(x)| \le G\|x' - x\|, \quad \forall x, x' \in \mathcal{S}.$$

Note that f may be non-differentiable. Assumption (2.17) is mainly used in the context of nonsmooth optimization [18, 89, 90].

Let $x^\star$ be an optimal point of the convex optimization problem (2.12) and use $f^\star$ to denote the corresponding optimal value. Given some $\varepsilon > 0$, assume that we want to estimate the number of iterations required by (2.16) to obtain ε-optimal solution, *i.e.*, a solution $x \in \mathcal{S}$ such that $f(x) - f^\star \le \varepsilon$. Using the step-size choice $\gamma = \varepsilon/G^2$, the mirror descent update (2.16) needs

$$T = \frac{2G^2 D_\omega\big(x(0), x^\star\big)}{\varepsilon^2} - 1$$

iterations to achieve ε-optimal solution [18]. This shows that under assumption (2.17), the iteration complexity of the the mirror descent method is $\mathcal{O}(1/\varepsilon^2)$. This complexity is optimal in the sense that it matches the lower complexity bound for convex optimization problems with Lipschitz continuous objective functions [87].

Objective functions with Lipschitz continuous gradients:
Next, we explore potential improvements in the convergence rate of the mirror descent method under certain smoothness assumptions. Suppose that f is continuously differentiable and, additionally, there is a constant $L \in (0, \infty)$ such that

$$\|\nabla f(x') - \nabla f(x)\|_* \le L\|x' - x\|, \quad x, x' \in \mathcal{S}. \tag{2.18}$$

An important consequence of (2.18) is that we can upper bound f everywhere by a quadratic function of fixed curvature:

$$f(x') \le f(x) + \langle \nabla f(x), x' - x \rangle + \frac{L}{2}\|x' - x\|^2, \quad \forall x, x' \in \mathcal{S}.$$

Under assumption (2.18), the sequence $\{x(k)\}$ generated by (2.16) with the constant step-size $\gamma = 1/L$ satisfies

$$f(x(k)) - f^\star \le \frac{LD_\omega\big(x(0), x^\star\big)}{k+1},$$

for all $k \in \mathbb{N}_0$ [91]. In view of this convergence rate, the time necessary to achieve ε-optimal solution is $\mathcal{O}(1/\varepsilon)$. This shows that using the assumption of Lipschitz continuity of the gradient, we are able to improve the iteration complexity of the mirror descent method (2.16) from $\mathcal{O}(1/\varepsilon^2)$ (in the nonsmooth case) to $\mathcal{O}(1/\varepsilon)$.

It is well known that for the class of convex optimization problems with Lipschitz continuous gradient, the optimal iteration complexity bound for finding ε-optimal solution is of the order $\mathcal{O}(1/\sqrt{\varepsilon})$ [85]. Therefore, the mirror descent method (2.16) is far from being optimal for smooth problems. In the seminal work [93], Nesterov presented a novel accelerated first-order method for unconstrained convex optimization that achieves the optimal iteration complexity $\mathcal{O}(1/\sqrt{\varepsilon})$. Some extensions of Nesterov's method were discussed in [94–97]. In particular, [96] proposed a fast gradient method for constrained convex problems using Bregman distance functions. That method generates iterates in the following manner

$$\begin{aligned} y(k) &= (1-\theta(k))z(k) + \theta(k)x(k), \\ x(k+1) &= \operatorname*{argmin}_{x\in\mathcal{S}} \Big\{ f(y(k)) + \big\langle \nabla f(y(k)), x - x(k) \big\rangle + \theta(k) L D_\omega\big(x(k), x\big) \Big\}, \\ z(k+1) &= (1-\theta(k))z(k) + \theta(k)x(k+1), \end{aligned} \tag{2.19}$$

with $x(0), z(0) \in \mathcal{S}$, $\theta(0) = 1$, and $\theta(k)$, $k \in \mathbb{N}$, given by

$$\theta(k+1) = \frac{\sqrt{\theta^4(k) + 4\theta^2(k)} - \theta^2(k)}{2}.$$

The sequence $\{z(k)\}$ produced by (2.19) satisfies

$$f(z(k)) - f^\star \le \frac{4LD_\omega\big(x(0), x^\star\big)}{(k+1)^2}, \quad k \in \mathbb{N}_0,$$

and hence $\mathcal{O}(1/\sqrt{\varepsilon})$ iterations suffice to obtain ε-optimal solution. Note that if $\theta(k) = 1$ for all $k \in \mathbb{N}$, then (2.19) reduces to the the mirror descent method (2.16). We can see that (2.19) is a variant of the mirror descent which is accelerated by two interpolation steps.

Chapter 3

Delay-independent Stability of Homogeneous Positive Systems

In distributed systems where exchange of information is involved, delays are inevitable. For this reason, a considerable effort has been devoted to characterizing the stability and performance of systems with delays (see, e.g., [98–102] and references therein). For general systems, the existence of time-delays may impair performance, induce oscillations and even instability [103]. In contrast, stability of positive linear systems is insensitive to certain classes of time-delays in the sense that if a positive linear system without delays is asymptotically stable, then it will remain asymptotically stable for any constant and bounded time-varying delays [104–109].

While the asymptotic stability of positive linear systems in the presence of time-delays has been thoroughly investigated, the theory for positive *nonlinear* systems is considerably less well-developed (see, e.g., [58, 80, 110] for exceptions). In particular, [80] showed that the asymptotic stability of a class of positive nonlinear systems whose vector fields are *cooperative* and *homogenous* of degree zero does not depend on the magnitude of *constant* delays. A similar result for cooperative systems that are homogeneous of any degree was given in [110], also under the assumption of constant delays. It is clear that considering constant delays is an idealized assumption as time-delays are often *time-varying* in practice. Stability analysis of positive nonlinear systems with time-varying delays is, however, challenging, since popular techniques for analyzing positive nonlinear systems with constant delays rely on a fundamental monotonicity property of trajectories of autonomous monotone systems, which does not hold when delays are time-varying [69].

At this point, it is worth noting that the results for positive linear systems cited above consider *bounded* delays. However, in some cases, it is not possible to guarantee a priori that the delays will be bounded. Instead, the state evolution might be affected by the entire history of states. It is then natural to ask if the insensitivity properties of positive linear systems with respect to time-delays will hold also for *unbounded* delays. In [111], it was shown that, for a particular class of unbounded delays, this is indeed the case. The question remains open for more general classes of time-delays. Moreover, [111] only considered the asymptotic stability and did not quantify how various bounds on the delay evolution impact the decay rate of positive linear systems.

Contributions of the Chapter. In this chapter, we consider time-varying, possibly unbounded, delays and establish the delay-independent stability of a class of positive nonlinear systems which includes positive linear systems as a special case. The proof technique uses neither the Lyapunov-Krasovskii functional method widely used to analyse positive linear systems with constant delays [104] nor the approach developed in [80, 110], and allows us to impose minimal restrictions on time-delays. Specifically, we make the following contributions:

- We derive a set of necessary and sufficient conditions for global delay independent stability of (i) continuous-time positive systems whose vector fields are cooperative and homogeneous of *arbitrary degree*, and (ii) discrete-time positive systems whose vector fields are order-preserving and homogeneous of *degree zero*. We demonstrate that the asymptotic stability of such systems is insensitive to a general class of time-delays which includes bounded and several types of unbounded time-varying delays as special cases.

- When the asymptotic behaviour of time-delays is known, we obtain conditions which ensure global μ-stability in the sense of [112]. These results allow us to give explicit estimates of the decay rate of homogeneous positive systems for various classes of (possibly unbounded) time-varying delays.

- For bounded delays and a particular class of unbounded delays, we present explicit expressions that quantify how the decay rate of homogeneous positive systems is affected by the upper bound of bounded time-varying delays, and the rate at which the unbounded delays grow large.

- For positive linear systems, we demonstrate that the best decay rate that our bounds can guarantee can be found via convex optimization.

- We also show that discrete-time positive systems whose vector fields are order-preserving and homogeneous of *degree greater than zero* are locally asymptotically stable under the global stability condition that we have derived for homogeneous positive systems of degree zero.

Outline of the Chapter. In Section 3.1, we define the class of continuous-time positive nonlinear systems for which we will study delay-independent stability. Then, we present our main results on global asymptotic stability and global μ-stability of such time-delay systems in Sections 3.2 and 3.3, respectively. In Section 3.4, we derive explicit expressions that allow us to quantify the impact of delays on the decay rate of positive linear systems. The corresponding counterparts for discrete-time positive nonlinear systems is given in Section 3.5. Illustrative examples are included throughout the development of the results. Finally, Section 3.6 summarizes the chapter. The appendix provides detailed proofs of the main results.

3.1 Problem Formulation

We consider the continuous-time dynamical system

$$\mathcal{G}: \quad \begin{cases} \dot{x}(t) = f\big(x(t)\big) + g\big(x(t-\tau(t))\big), & t \geq 0, \\ x(t) = \varphi(t), & t \in [-\tau_{\max}, 0], \end{cases} \tag{3.1}$$

where $x(t) \in \mathbb{R}^n$ is the state variable, $f, g : \mathbb{R}^n \to \mathbb{R}^n$ are continuously differentiable vector fields with $f(\mathbf{0}) = g(\mathbf{0}) = \mathbf{0}$, $\tau_{\max} \in \mathbb{R}_+$, and $\varphi \in \mathcal{C}\big([-\tau_{\max}, 0], \mathbb{R}^n\big)$ is the vector-valued function specifying the initial condition of the system. Unlike non-delayed systems, the solution of the time-delay system $\mathcal{G}$ given by (3.1) is not uniquely determined by a point-wise initial condition $x(0)$, but by the continuous function φ defined over the interval $[-\tau_{\max}, 0]$. We assume that the time-varying delay $\tau(t)$ satisfies the following assumption:

Assumption 3.1 ([**16, §6.1**]). *The delay $\tau : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous with respect to time and satisfies*

$$\lim_{t \to +\infty} t - \tau(t) = +\infty. \tag{3.2}$$

Note that $\tau(t)$ is not necessarily continuously differentiable and no restriction on its derivative (such as $\dot{\tau}(t) < 1$) is imposed. Roughly speaking, condition (3.2) implies that as t increases, $\tau(t)$ grows slower than time itself. This constraint on time-delays is typically satisfied in real-world applications. For example, a class of continuous-time power control algorithms for a wireless network consisting of n mobile users can be described by

$$\dot{x}_i\big(t\big) = -x_i\big(t\big) + \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j\big(t - \tau(t)\big), \quad i = 1, \dots, n. \tag{3.3}$$

Here, $x_i(t)$ is the transmitted power of user i at time t, and $a_{ij} \in \mathbb{R}_+$, $i, j = 1, \dots, n$, are nonnegative constants [113]. If the delay $\tau(t)$ satisfies condition (3.2), then given any time $t_1 \geq 0$, there exists a time $t_2 \geq t_1$ such that

$$t - \tau(t) \geq t_1, \quad \forall t \geq t_2.$$

This simply means that given any time t_1, information about which transmitted power each user has applied prior to t_1 will be received by every other user before a sufficiently long time t_2 and not be used in the state evolution of (3.3) after t_2. In other words, state information eventually propagates to all other users in the network and old information is eventually purged from the network. Therefore, in the power control problem, Assumption 3.1 is always satisfied unless the communication between users is totally lost during a semi-infinite time interval.

Note that all bounded delays, irrespectively of whether they are constant or time-varying, satisfy Assumption 3.1. Moreover, delays satisfying (3.2) may be unbounded. Consider the following particular class of unbounded delays which was studied in [111, 114–116].

Assumption 3.2. *There exist $T > 0$ and a scalar $\alpha \in [0,1)$ such that*

$$\sup_{t \geq T} \frac{\tau(t)}{t} = \alpha. \tag{3.4}$$

One can easily verify that constraint (3.4) on time-delays implies (3.2). However, the next example shows that the converse does, in general, not hold. Hence, Assumption 3.2 is a special case of Assumption 3.1.

Example 3.1. Let $\tau(t) = t - \ln(t+1)$, $t \geq 0$. Since

$$\begin{aligned} \lim_{t\to+\infty} t - \tau(t) &= \lim_{t\to+\infty} \ln(t+1) = +\infty, \\ \lim_{t\to+\infty} \frac{\tau(t)}{t} &= \lim_{t\to+\infty} \frac{t - \ln(t+1)}{t} = 1, \end{aligned}$$

it is clear that (3.2) holds while (3.4) does not hold. ■

Remark 3.1. Assumption 3.1 implies that there exists a sufficiently large $T_0 > 0$ such that $t - \tau(t) > 0$ for all $t > T_0$. Define

$$\tau_{\max} = - \inf_{0 \leq t \leq T_0} \Big\{ t - \tau(t) \Big\}.$$

Since $\tau_{\max} \in \mathbb{R}_+$ is bounded ($\tau_{\max} < +\infty$), it follows that for any delay satisfying Assumption 3.1, even if it is unbounded, the initial condition φ is defined on a bounded set $[-\tau_{\max}, 0]$.

In this chapter, we study delay-independent stability of nonlinear systems of the form (3.1) which are *positive* defined as follows.

Definition 3.1. The time-delay system $\mathcal{G}$ given by (3.1) is said to be positive if for every nonnegative initial condition $\varphi \in \mathcal{C}\big([-\tau_{\max}, 0], \mathbb{R}^n_+\big)$, the corresponding state trajectory is nonnegative, that is $x(t) \in \mathbb{R}^n_+$ for all $t \geq 0$.

The following result provides a sufficient condition for positivity of dynamical systems with time-varying delays.

Proposition 3.1 ([**69, Theorem 5.2.1**]). *For the time-delay system $\mathcal{G}$ given by* (3.1)*, suppose that*

$$\begin{aligned} f_i(x) &\geq 0, \quad \text{for all } i = 1, \ldots, n, \text{ and } x \in \mathbb{R}^n_+ \text{ such that } x_i = 0, \\ g(x) &\geq \mathbf{0}, \quad \text{for all } x \in \mathbb{R}^n_+. \end{aligned} \tag{3.5}$$

Then, $\mathcal{G}$ is positive.

Proof. See Appendix 3.7.1. ■

The following example illustrates the result of Proposition 3.1.

Example 3.2. Consider the time-delay system $\mathcal{G}$ given by (3.1) with

$$f(x_1, x_2) = \begin{bmatrix} -x_1^2 + 4x_2 - 3x_1^2 x_2^2 \\ 2x_1 - 4x_2^2 \end{bmatrix}, \quad g(x_1, x_2) = \begin{bmatrix} x_1 x_2 \\ x_1^2 + x_2 \end{bmatrix}.$$

If $x = (x_1, x_2) \in \mathbb{R}_+^2$, we then have

$$\begin{aligned} f_1(x_1 = 0, x_2) &= 4x_2 \geq 0, \\ f_2(x_1, x_2 = 0) &= 2x_1 \geq 0, \end{aligned}$$

and $g(x_1, x_2) \geq (0, 0)$. Therefore, according to Proposition 3.1, this system is positive. ■

Note that the nonnegativity of initial conditions is essential for ensuring positivity of the state evolution of $\mathcal{G}$. In other words, when $\varphi(t) \geq \mathbf{0}$, $t \in [-\tau_{\max}, 0]$, is not satisfied, $x(t)$ may not stay in the positive orthant even if (3.5) holds.

In [58, Proposition 3.1], it was shown that when delays are *constant*, the sufficient condition given in Proposition 3.1 is also necessary for positivity of the time-delay system $\mathcal{G}$ given by (3.1), *i.e.*, $\mathcal{G}$ with $\tau(t) = \tau_{\max} > 0$, $t \geq 0$, is positive if and only if (3.5) holds. However, as the next example shows, this condition is not necessary when we allow for *time-varying* delays.

Example 3.3. Consider a continuous-time linear system described by (3.1) with

$$f(x_1, x_2) = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad g(x_1, x_2) = \begin{bmatrix} 0 & 0 \\ e & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{3.6}$$

where e is the base of the natural logarithm, and let the time-varying delay be

$$\tau(t) = \begin{cases} 0, & 0 \leq t \leq 1, \\ t - 1, & 1 \leq t \leq 2, \\ 1, & 2 \leq t. \end{cases} \tag{3.7}$$

The solution to the time-delay system (3.6) is given by

$$\begin{aligned} x_1(t) &= x_1(0)e^t, \qquad\qquad 0 \leq t, \\ x_2(t) &= \begin{cases} x_2(0) + (e-1)(e^t - 1)x_1(0), & 0 \leq t \leq 1, \\ x_2(0) + (e^2 t - e^t + 1 - e)x_1(0), & 1 \leq t \leq 2, \\ x_2(0) + (e^2 - e + 1)x_1(0), & 2 \leq t. \end{cases} \end{aligned}$$

It is straightforward to verify that if $x(0) = (x_1(0), x_2(0)) \in \mathbb{R}_+^2$, then

$$x(t) = (x_1(t), x_2(t)) \in \mathbb{R}_+^2,$$

for all $t \geq 0$. Therefore, the linear system (3.6) with the bounded time-varying delay (3.7) is positive. However, the sufficient condition given in Proposition 3.1 is not satisfied in this example since $f_2(x_1, x_2) \geq 0$ does not hold for all $x \in \mathbb{R}_+^2$ with $x_2 = 0$ (take, for example, $f_2(1, 0) = -1 < 0$). ■

From this point on, we assume that f and g satisfy Assumption 3.3.

Assumption 3.3. *The following properties hold:*

1. *f is cooperative and g is order-preserving on $\mathbb{R}^n_+$.*

2. *f and g are homogeneous of degree p with respect to the dilation map $\delta^r_\lambda(x)$.*

A system $\mathcal{G}$ given by (3.1) satisfying Assumption 3.3 is called *homogeneous cooperative*. Since $f(\mathbf{0}) = g(\mathbf{0}) = \mathbf{0}$, it follows from Proposition 3.1 that Assumption 3.3.1 ensures the positivity of homogeneous cooperative systems. The model of some physical systems falls within this class of positive systems. For example, continuous-time linear and several nonlinear power control algorithms for wireless networks are described by homogeneous cooperative systems [117].

While the presence of time-delays typically affects the stability of general dynamical systems, the global asymptotic stability of homogeneous cooperative systems is independent of constant delays [110]. More precisely, the homogeneous cooperative system (3.1) with a constant delay $\tau(t) = \tau_{\max}$, $t \geq 0$, is globally asymptotically stable for all $\tau_{\max} > 0$ if and only if the corresponding delay-free system ($\tau_{\max} = 0$) is globally asymptotically stable. Our main objectives are therefore to (i) determine whether a similar delay-independent stability result holds for homogeneous cooperative systems with time-varying delays satisfying Assumption 3.1; and to (ii) give explicit estimates of the decay rate for different classes of time-delays (e.g., bounded delays, unbounded delays satisfying Assumption 3.2, etc.).

3.2 Asymptotic Stability of Homogeneous Cooperative Systems

The following theorem establishes a necessary and sufficient condition for global asymptotic stability of homogeneous cooperative systems with time-varying delays satisfying Assumption 3.1. Our proof (which is similar to that in [16] for asynchronous *discrete-time* systems and conceptually related to the Lyapunov stability theorem) uses the Lyapunov function

$$V(x) = \max_{1\leq i\leq n}\left(\frac{x_i}{v_i}\right)^{\frac{r_{\max}}{r_i}}, \tag{3.8}$$

where r_i, $i = 1, \dots, n$, are defined by the dilation map $\delta^r_\lambda(x)$, $r_{\max} = \max_{1\leq i\leq n} r_i$, and $v = (v_1, \dots, v_n) \in \mathbb{R}^n$ is a positive vector ($v > \mathbf{0}$). We define sets

$$S(m) := \left\{x \in \mathbb{R}^n_+ \,\middle|\, V(x) \leq \gamma^m \|\varphi\|\right\}, \quad m \in \mathbb{N}_0, \tag{3.9}$$

where $\gamma \in [0, 1)$, and

$$\|\varphi\| := \sup_{-\tau_{\max}\leq s\leq 0} V(\varphi(s)). \tag{3.10}$$

Then, we show that for each m, there exists $t_m \geq 0$ such that $x(t) \in S(m)$ for all $t \geq t_m$. In other words, the system state starting from nonnegative initial conditions will enter each set $S(m)$ at some time t_m and remain in the set for all future times. Since the sets are nested, *i.e.*,

$$S(0) \supset \cdots \supset S(m) \supset S(m+1) \supset \cdots,$$

the state will move sequentially from set $S(m)$ to $S(m+1)$, cf. Figure 3.1. Thus,

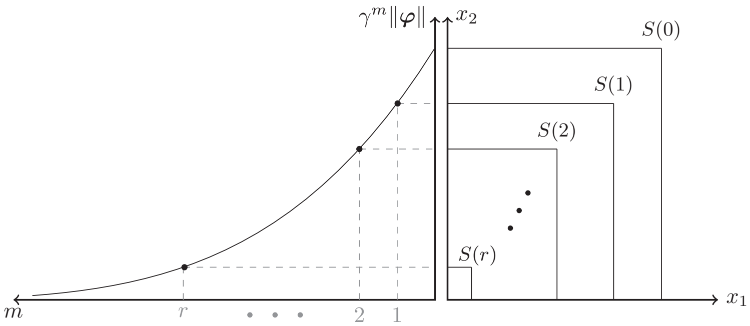


Figure 3.1: Level curves of the Lyapunov function $V(x)$ in the two-dimensional case. The key idea behind the proof of Theorem 3.1 is that $\varphi(t)$, $t \in [-\tau_{\max}, 0]$, is initially within the set $S(0)$. At some time $t_1 \geq 0$, the system state $x(t)$ eventually enters and stays within the set $S(1)$ for all $t \geq t_1$; moreover, as t increases further, $x(t)$ sequentially moves into other sets.

the sets play a similar role as level sets of the Lyapunov function $V(x)$. Note that when f and g are homogeneous with respect to the standard dilation map ($r_i = 1$ for each i), the Lyapunov function (3.8) reduces to $V(x) = \|x\|_\infty^v$, which is often used in stability analysis of positive linear systems [64].

Theorem 3.1. *For the time-delay system $\mathcal{G}$ given by* (3.1)*, suppose that Assumptions 3.1 and 3.3 hold. Then, the following statements are equivalent.*

(i) There exists a vector $v > \mathbf{0}$ such that

$$f(v) + g(v) < \mathbf{0}. \tag{3.11}$$

(ii) The homogeneous cooperative system $\mathcal{G}$ is globally asymptotically stable for every nonnegative initial condition $\varphi \in \mathcal{C}([-\tau_{\max}, 0], \mathbb{R}_+^n)$, and for all time-delays satisfying Assumption 3.1.

(iii) The homogeneous cooperative system $\mathcal{G}$ without delay ($\tau(t) = 0,\ t \geq 0$) is globally asymptotically stable for all nonnegative initial conditions.

Proof. See Appendix 3.7.2. ∎

According to Theorem 3.1, if we can demonstrate the existence of a vector $v > \mathbf{0}$ satisfying (3.11), then the homogeneous cooperative system $\mathcal{G}$ given by (3.1) is globally asymptotically stable for all time-delays satisfying Assumption 3.1. In other words, the global asymptotic stability does not depend on the magnitude and variation of the delays, but only on the vector fields. This property is very useful in practical applications since the delays may not be easy to model in detail.

We now present a simple example to illustrate the use of Theorem 3.1.

Example 3.4. Consider the time-delay dynamical system $\mathcal{G}$ given by (3.1) with

$$f(x_1, x_2) = \begin{bmatrix} -5x_1^3 + x_1 x_2^4 \\ x_1^2 x_2 - 5x_2^5 \end{bmatrix}, \quad g(x_1, x_2) = \begin{bmatrix} 2x_1^2 x_2^2 \\ x_1 x_2^3 \end{bmatrix}. \tag{3.12}$$

Computing the Jacobian matrix $\partial f / \partial x$ at a point (x_1, x_2), we have

$$\frac{\partial f}{\partial x}(x_1, x_2) = \begin{bmatrix} -15x_1^2 + x_2^4 & 4x_1 x_2^3 \\ 2x_1 x_2 & x_1^2 - 25x_2^4 \end{bmatrix}$$

For all $(x_1, x_2) \in \mathbb{R}_+^2$, the off-diagonal entries of this matrix are nonnegative, *i.e.*, $\partial f / \partial x$ is Metzler. Therefore, f is cooperative on $\mathbb{R}_+^2$. Let $\delta_\lambda^r(x)$ be the dilation map with $r = (2, 1)$. Since

$$f\big(\delta_\lambda^r(x)\big) = f\big(\lambda^2 x_1, \lambda x_2\big) = \lambda^4 \begin{bmatrix} \lambda^2(-5x_1^3 + x_1 x_2^4) \\ \lambda(x_1^2 x_2 - 5x_2^5) \end{bmatrix} = \lambda^4 \delta_\lambda^r\big(f(x)\big),$$

f is homogeneous of degree $p = 4$ with respect to $\delta_\lambda^r(x)$. It is straightforward to verify that g is order-preserving on $\mathbb{R}_+^2$ and homogeneous of degree $p = 4$ with respect to the dilation map $\delta_\lambda^r(x)$ with $r = (2, 1)$. Since $f(1,1) + g(1,1) = (-2, -3) < (0, 0)$, it follows from Theorem 3.1 that for any nonnegative initial conditions, the homogeneous cooperative system (3.12) without time-delays and with time-delays satisfying Assumption 3.1 is globally asymptotically stable. ■

Note that Theorem 3.1 can be easily extended to homogeneous cooperative systems with multiple delays of the form

$$\dot{x}\big(t\big) = f\big(x(t)\big) + \sum_{j=1}^{J} g_j\big(x(t - \tau_j(t))\big).$$

Here, f is cooperative and homogeneous, g_j for $j = 1, \ldots, J$ are homogenous and order-preserving on $\mathbb{R}_+^n$, and $\tau_j(t)$ satisfy Assumption 3.1. In this case, the stability condition (3.11) becomes

$$f(v) + \sum_{j=1}^{J} g_j(v) < \mathbf{0},$$

for some $v > \mathbf{0}$.

3.3 Decay Rates of Homogeneous Cooperative Systems

Theorem 3.1 is concerned with the *asymptotic* stability of homogeneous cooperative systems with time-varying delays. However, there are processes and applications for which it is desirable that the system has a certain decay rate. Loosely speaking, the system has to converge quickly enough to the equilibrium. It turns out quantitative stability measures, such as the decay rate, can be highly dependent on the magnitude of delays. In this section, we therefore characterize how time-delays affect the decay rate of homogeneous cooperative systems. Before stating the main result, we provide the definition of μ-stability, introduced in [112], for continuous-time systems.

Definition 3.2. Suppose that $\mu : \mathbb{R}_+ \to \mathbb{R}_+$ is a non-decreasing function satisfying $\mu(t) \to +\infty$ as $t \to +\infty$. The time-delay system $\mathcal{G}$ given by (3.1) is said to be globally μ-stable if for any initial function φ, there exists a constant $M > 0$ such that the solution $x(t)$ satisfies

$$\|x(t)\| \le \frac{M}{\mu(t)}, \quad t > 0,$$

where $\|\cdot\|$ is some norm on $\mathbb{R}^n$.

This definition can be regarded as a unification of several types of stability. For example, when $\mu(t) = e^{\eta t}$ with $\eta > 0$, μ-stability becomes *exponential stability*; and when $\mu(t) = t^{\xi}$ with $\xi > 0$, then μ-stability becomes *power-rate stability.*

Global μ-stability of homogenous cooperative systems with bounded and unbounded time-varying delays can be verified using the following theorem.

Theorem 3.2. *Consider the time-delay system $\mathcal{G}$ given by* (3.1). *Suppose that Assumptions 3.1 and 3.3 hold, and that there is a vector $v > \mathbf{0}$ satisfying*

$$f(v) + g(v) < \mathbf{0}. \tag{3.13}$$

Assume also that there exists a function $\mu : \mathbb{R}_+ \to \mathbb{R}_+$ such that the following conditions hold:

(i) $\mu(t) > 0$, *for all* $t > 0$.

(ii) $\mu(t)$ *is non-decreasing.*

(iii) $\lim_{t\to+\infty} \mu(t) = +\infty$.

(iv) For each $i \in \{1, \ldots, n\}$,

$$\left(\frac{r_{\max}}{r_i}\right)\left(\left(\frac{f_i(v)}{v_i}\right) + \left(\lim_{t\to\infty} \frac{\mu(t)}{\mu(t-\tau(t))}\right)^{\frac{r_i+p}{r_{\max}}}\left(\frac{g_i(v)}{v_i}\right)\right) + \lim_{t\to\infty} \frac{\dot{\mu}(t)}{(\mu(t))^{1-\frac{p}{r_{\max}}}} < 0.$$

Then, every solution of $\mathcal{G}$ starting in the positive orthant satisfies

$$\left(\frac{x_i(t)}{v_i}\right)^{\frac{r_{\max}}{r_i}} = O\left(\mu^{-1}(t)\right), \quad t \ge 0,$$

for $i = 1, \ldots, n$.

Proof. See Appendix 3.7.3. ∎

According to Theorem 3.2, any function $\mu(t)$ satisfying conditions (i)–(iv) can be used to estimate the decay rate of homogeneous cooperative systems with time-varying delays. From condition (iv), it is clear that the asymptotic behaviour of the delay $\tau(t)$ influences the admissible choices for $\mu(t)$ and, hence, the decay bounds that we are able to guarantee. To clarify this statement, we will analyze a few special cases in detail. First, assume that $\tau(t)$ is bounded, *i.e.*,

$$0 \leq \tau(t) \leq \tau_{\max}, \quad t \geq 0. \tag{3.14}$$

The following result shows that under this assumption, the decay rate of homogeneous cooperative systems of degree p is upper bounded by an exponential function of time when $p = 0$ and by a polynomial function of time when $p > 0$.

Corollary 3.1. *For the homogeneous cooperative system* $\mathcal{G}$ *given by* (3.1)*, suppose that there exists a vector* $v > \mathbf{0}$ *satisfying* (3.13)*, and that* (3.14) *holds.*

(i) If f *and* g *are homogeneous of degree zero, then* $\mathcal{G}$ *is globally exponentially stable. In particular,*

$$\left(\frac{x_i(t)}{v_i}\right)^{\frac{r_{\max}}{r_i}} = O\left(e^{-\eta t}\right), \quad t \geq 0, \tag{3.15}$$

where $\eta \in (0, \min_{1\leq i\leq n} \eta_i)$*, and* η_i *is the positive solution of the equation*

$$\left(\frac{r_{\max}}{r_i}\right)\left(\left(\frac{f_i(v)}{v_i}\right) + \left(e^{\eta_i \tau_{\max}}\right)^{\frac{r_i}{r_{\max}}}\left(\frac{g_i(v)}{v_i}\right)\right) + \eta_i = 0. \tag{3.16}$$

(ii) If f *and* g *are homogeneous of degree greater than zero, then*

$$\left(\frac{x_i(t)}{v_i}\right)^{\frac{r_{\max}}{r_i}} = O\left((\theta t + 1)^{\frac{-r_{\max}}{p}}\right), \quad t \geq 0, \tag{3.17}$$

where

$$\theta \in \left(0, \min\left\{\frac{1}{\tau_{\max}}, \min_{1\leq i\leq n} \theta_i\right\}\right),$$

and θ_i *is the positive solution to*

$$\frac{f_i(v)}{v_i} + \frac{g_i(v)}{v_i} + \theta_i \frac{r_i}{p} = 0. \tag{3.18}$$

Proof. See Appendix 3.7.4. ∎

According to Corollary 3.1, if the delay-independent stability condition (3.13) holds, homogeneous cooperative systems of degree zero with bounded delays are globally exponentially stable. In addition, any vector $v > \mathbf{0}$ satisfying (3.13) can be used to find a guaranteed decay rate by computing the associated η in (3.16). Equation (3.16) has three parameters: the maximum delay bound $\tau_{\max}$, the positive vector v, and η_i. For any fixed $\tau_{\max} \in \mathbb{R}_+$ and any fixed $v > \mathbf{0}$ satisfying (3.13), the left-hand side of (3.16) is smaller than the right-hand side for $\eta_i = 0$, and strictly monotonically increasing in $\eta_i > 0$. Therefore, (3.16) has always a unique positive solution η_i. Note that η_i is monotonically decreasing in $\tau_{\max}$ and approaches zero as $\tau_{\max}$ tends to infinity. Hence, the guaranteed decay rate slows down as the delays increase in magnitude.

The following example illustrates how Corollary 3.1 can help us to obtain an estimate of the decay rate of homogeneous cooperative systems with bounded time-varying delays.

Example 3.5. Consider the time-delay system $\mathcal{G}$ given by (3.1) with

$$f(x_1, x_2) = \begin{bmatrix} -5x_1^3 + 2x_1x_2 \\ x_1^2x_2 - 4x_2^2 \end{bmatrix}, \quad g(x_1, x_2) = \begin{bmatrix} x_1x_2 \\ 2x_1^4 \end{bmatrix}. \tag{3.19}$$

Both f and g are homogeneous of degree $p = 2$ with respect to the dilation map $\delta_\lambda^r(x)$ with $r = (1, 2)$. Moreover, f is cooperative and g is order-preserving on $\mathbb{R}_+^2$. Since $f(1, 1) + g(1, 1) = (-2, -1) < (0, 0)$, according to Corollary 3.1, the decay rate of the homogeneous cooperative system (3.19) with bounded delays is upper bounded by a polynomial function of time. Now, consider the specific time-delay $\tau(t) = 4 + \sin(t)$, $t \geq 0$. Clearly, $\tau(t) \leq \tau_{\max} = 5$ for all $t \geq 0$. Using $v = (1, 1)$ and $(r_1, r_2) = (1, 2)$, the solutions to (3.18) are $\theta_1 = 4$ and $\theta_2 = 1$, which implies that

$$\theta \in \left(0, \min\left\{\frac{1}{5}, \min\{4, 1\}\right\} = \frac{1}{5}\right).$$

Thus, for nonnegative initial conditions, the system state $x(t)$ satisfies

$$\max\{x_1^2(t), x_2(t)\} = O\left(\frac{1}{\frac{1}{5}t + 1}\right), \quad t \geq 0.$$

Figure 3.2 gives the simulation results of the actual decay rate of the homogeneous cooperative system (3.19) and the guaranteed decay rate we calculated, when the initial condition is $\varphi(t) = (1, 1)$, $t \in [-5, 0]$. ■

While the stability of homogeneous cooperative systems with delays satisfying Assumption 3.1 may, in general, only be asymptotic, Corollary 3.1 demonstrates that if delays are bounded, we can guarantee certain decay rates. We will now establish similar decay bounds for unbounded delays satisfying Assumption 3.2.

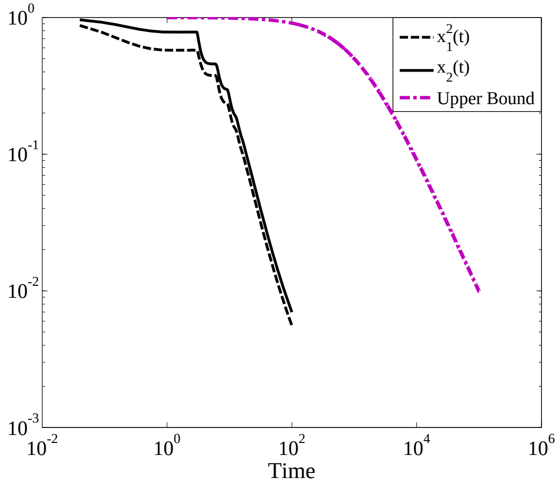


Figure 3.2: Comparison of guaranteed upper bound and actual decay rate of the homogeneous cooperative system (3.19) with bounded time-varying delays.

Corollary 3.2. *Consider the homogeneous cooperative system $\mathcal{G}$ given by* (3.1)*. If there exists a vector $v > \mathbf{0}$ such that* (3.13) *holds, then the positive system $\mathcal{G}$ with delays satisfying Assumption 3.2 is globally power-rate stable. In particular,*

(i) if f and g are homogeneous of degree zero, the solution $x(t)$ of $\mathcal{G}$ satisfies

$$\left(\frac{x_i(t)}{v_i}\right)^{\frac{r_{\max}}{r_i}} = O\left(t^{-\xi}\right), \quad t \geq 0,$$

where $\xi \in (0, \min_{1\leq i\leq n} \xi_i)$, and ξ_i is the unique positive solution to

$$\left(\frac{f_i(v)}{v_i}\right) + \left(\frac{1}{1-\alpha}\right)^{\frac{r_i}{r_{\max}}\xi_i} \left(\frac{g_i(v)}{v_i}\right) = 0 \tag{3.20}$$

(ii) if f and g are homogeneous of degree greater than zero, then

$$\left(\frac{x_i(t)}{v_i}\right)^{\frac{r_{\max}}{r_i}} = O\left(t^{\frac{-r_{\max}}{p}\beta}\right), \quad t \geq 0,$$

where $\beta \in (0, 1)$ is such that

$$\left(\frac{f_i(v)}{v_i}\right) + \left(\frac{1}{1-\alpha}\right)^{(1+\frac{r_i}{p})\beta} \left(\frac{g_i(v)}{v_i}\right) < 0, \tag{3.21}$$

holds for each i.

Proof. See Appendix 3.7.5. ■

Corollary 3.2 shows that the decay rate of homogeneous cooperative systems of degree zero with unbounded delays satisfying Assumption 3.2 is of order $O(t^{-\xi})$. Equation (3.20) quantifies how α, the rate at which the unbounded delays grow large, affects ξ. Specifically, ξ_i is monotonically decreasing with α and approaches zero as α tends to one. From (3.21), we see that β, on which the guaranteed decay rate of homogeneous cooperative systems of degree greater than zero depends, also approaches zero as α tends to one (see Figure 3.3). Hence, the guaranteed convergence rates of homogeneous cooperative systems with unbounded delays satisfying Assumption 3.2 deteriorate with increasing α.

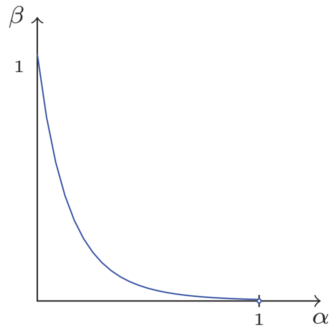


Figure 3.3: Plot of β in (3.21) for different values of $\alpha \in [0, 1)$. Clearly, β is monotonically decreasing with α and approaches zero as α tends to one.

3.4 A Special Case: Positive Linear Systems

We now discuss the delay-independent stability of a special case of (3.1), namely the continuous-time linear system $\mathcal{G}_L$ of the form

$$\mathcal{G}_L : \begin{cases} \dot{x}(t) = Ax(t) + Bx(t - \tau(t)), & t \geq 0, \\ x(t) = \varphi(t), & t \in [-\tau_{\max}, 0]. \end{cases} \tag{3.22}$$

In terms of (3.1), $f(x) = Ax$ and $g(x) = Bx$. If $A \in \mathbb{R}^{n\times n}$ is Metzler and $B \in \mathbb{R}^{n\times n}$ is nonnegative, then Assumption 3.3 is satisfied, which means the time-delay linear system $\mathcal{G}_L$ is homogeneous cooperative (of degree zero) and, hence, positive.

We have the following special case of Theorem 3.1.

Corollary 3.3. *Consider the time-delay linear system $\mathcal{G}_L$ given by* (3.22)*. Assume that A is Metzler and B is nonnegative. Then, the positive linear system $\mathcal{G}_L$ is* *globally asymptotically stable for all time-delays satisfying Assumption 3.1 if and only if there exists a vector v such that*

$$\begin{cases} (A+B)v < \mathbf{0}, \\ \qquad\quad v \;\; > \mathbf{0}. \end{cases} \tag{3.23}$$

The stability condition (3.23) is a linear programming problem in v, and thus can be verified numerically in polynomial time [118]. Since A is Metzler and B is nonnegative, $A+B$ is Metzler. It follows from Theorem 2.3 that (3.23) has a feasible solution v if and only if $A+B$ is Hurwitz, *i.e.*, all its eigenvalues have negative real parts. Therefore, according to Corollary 3.3, if the positive linear system (3.22) without delay is stable ($A+B$ is Hurwitz), it remains asymptotically stable under all bounded and unbounded time-varying delays satisfying Assumption 3.1.

While the asymptotic stability of positive linear systems with delays satisfying Assumption 3.1 has been investigated in [119], the impact of time-delays on the decay rate has not been analyzed. Theorem 3.2 can help us to find guaranteed decay rates of $\mathcal{G}_L$ for different classes of time delays. Specifically, Corollaries 3.1 and 3.2 show that $\mathcal{G}_L$ is exponentially stable if the time-delays are bounded, and power-rate stable if the unbounded delays satisfy Assumption 3.2. Therefore, not only do we extend the result of [119] to general homogeneous cooperative (not necessarily linear) systems, but we also provide explicit bounds on the decay rate of positive linear systems.

The associated convergence rate result for positive linear systems with time-varying delays reads as follows:

Corollary 3.4. *For the time-delay linear system $\mathcal{G}_L$ given by* (3.22)*, assume that A is Metzler and B is nonnegative. In addition, assume that there is a vector v satisfying* (3.23)*. Then, the following statements hold:*

(i) For any bounded delays satisfying (3.14)*, $\mathcal{G}_L$ is globally exponentially stable. In particular, the solution $x(t)$ of $\mathcal{G}_L$ for nonnegative initial conditions satisfies*

$$\|x(t)\|_\infty^v = O\left(e^{-\eta t}\right), \quad t \geq 0,$$

where $\eta \in (0, \min_{1\leq i\leq n} \eta_i)$, and η_i is the positive solution of the equation

$$\left(\sum_{j=1}^n \frac{1}{v_i} a_{ij} v_j\right) + \left(\sum_{j=1}^n \frac{1}{v_i} b_{ij} v_j\right) e^{\eta_i \tau_{\max}} + \eta_i = 0. \tag{3.24}$$

(ii) For any unbounded delays satisfying Assumption 3.2, we have

$$\|x(t)\|_\infty^v = O\left(t^{-\xi}\right), \quad t \geq 0,$$

where $\xi \in (0, \min_{1\leq i\leq n} \xi_i)$, and ξ_i is the unique positive solution to

$$\left(\sum_{j=1}^n \frac{1}{v_i} a_{ij} v_j\right) + \left(\sum_{j=1}^n \frac{1}{v_i} b_{ij} v_j\right)\left(\frac{1}{1-\alpha}\right)^{\xi_i} = 0. \tag{3.25}$$

As shown in Corollary 3.4, when time-delays are bounded, any feasible solution v to the linear programming problem (3.23) can be used to find a guaranteed exponential bound on the decay rate of positive linear systems by computing the associated η. From (3.24), it is easily seen that η depends on the choice of vector v, and that an arbitrary feasible v not necessarily gives a tight guaranteed bound on the actual decay rate. Next, we will show that the best decay rate that our results can ensure, along with the associated vector v can be found via convex optimization. To this end, we use the logarithmic change of variables $z_i = \ln(v_i)$, $i = 1, \ldots, n$. This change of variables is valid since v_i is required to be positive for each i. The search for the best guaranteed decay rate can be formulated as

$$\begin{aligned} \textbf{maximize} \quad & \eta \\ \textbf{subject to} \quad & \eta < \eta_i, && (3.26a) \\ & a_{ii} + b_{ii} + \sum_{j \neq i} \big(a_{ij} + b_{ij} \big) e^{z_j - z_i} < 0, && (3.26b) \\ & a_{ii} + \sum_{j \neq i} a_{ij} e^{z_j - z_i} + \sum_{j=1}^{n} b_{ij} e^{z_j - z_i + \eta_i \tau_{\max}} + \eta_i \leq 0, && (3.26c) \end{aligned}$$

where the last two constraints are (3.23) and (3.24) in the new variables, respectively. The optimization variables are the decay rate η and the vector $z = (z_1, \ldots, z_n)$. Since $a_{ij} \geq 0$ for all $i \neq j$ and $b_{ij} \geq 0$ for all $i, j = 1, \ldots, n$, the last two constraints in (3.26) are convex in η and z. This implies that (3.26) is a convex optimization problem; hence, it can be efficiently solved [84]. Note that if the delays satisfy Assumption (3.2), we can solve again the convex optimization problem (3.26) to find the best guaranteed decay rate. The only change is that we replace η with ξ, and the last constraint with

$$a_{ii} + \sum_{j \neq i} a_{ij} e^{z_j - z_i} + \sum_{j=1}^{n} b_{ij} e^{z_j - z_i + (\ln \frac{1}{1-\alpha}) \xi_i} \leq 0.$$

We now provide an example.

Example 3.6. Consider the time-delay linear system (3.22) with

$$A = \begin{bmatrix} -6 & 2 \\ 1 & -3 \end{bmatrix}, \; B = \begin{bmatrix} 3 & 0 \\ 0 & 0.5 \end{bmatrix}. \tag{3.27}$$

Let the time-varying delay be given by $\tau(t) = 3 + \cos(t)$, $t \geq 0$. Obviously, one may choose $\tau_{\max} = 4$ as an upper bound on this delay. The matrix A is Metzler and B is nonnegative. Thus, the time-delay linear system (3.27) is positive. The eigenvalues of $A + B$ are -4.18 and -1.31. Since $A + B$ is Hurwitz, it follows from Corollary 3.4 that the positive system (3.27) with bounded time-varying delays is globally exponentially stable. We will now calculate guaranteed exponential upper bounds on the decay rate.

Since $A+B$ is Hurwitz, the following linear programming problem admits a solution

$$\begin{cases} \begin{bmatrix} -3 & 2 \\ 1 & -2.5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} < \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \\ \qquad\qquad\qquad \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} > \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{cases} \tag{3.28}$$

As discussed before, any feasible solution v to (3.28) yields a guaranteed decay rate by computing the associated η in (3.24). For example, if we take $v^1 = (1,1)$, the solutions to the nonlinear equation (3.24) can be obtained as

$$\eta_1 = 0.0677, \ \eta_2 = 0.3052,$$

implying that $\eta = \min\{0.0677, 0.3052\} = 0.0677$. Thus, for nonnegative initial conditions, we have

$$\|x(t)\|_\infty^{v^1} = O\left(e^{-0.0677t}\right), \quad t \geq 0.$$

The left-hand side of Figure 3.4 compares $\|x(t)\|_\infty^{v^1}$ obtained by simulating (3.27) from the initial condition $\varphi(t) = v^1$, $t \in [-4, 0]$, and the theoretical upper bound e^{-0677t}. Of course, v^1 is only one of the possible solutions of (3.28). Next, by solving the convex optimization problem (3.26), we get

$$v^\star = (0.9020, \ 0.4317), \ \eta^\star = 0.0838,$$

which implies that the solution $x(t)$ satisfies

$$\|x(t)\|_\infty^{v^\star} = O\left(e^{-0.0838t}\right), \quad t \geq 0.$$

The right-hand side of Figure 3.4 gives the simulation results of $\|x(t)\|_\infty^{v^\star}$, and the theoretical upper bound $e^{-0.0838t}$ when the initial condition is $\varphi(t) = v^\star$, $t \in [-4, 0]$. We can see that the linear inequalities (3.28) do not help us in guiding our search for a vector v which guarantees a fast decay rate. In contrast, solving the convex optimization problem (3.26) finds the best $\eta^\star$ that our bound can guarantee along with the associated $v^\star$. The bound matches simulations very well and is a significant improvement over simply using the non-optimized v^1. ■

Remark 3.2. In [120, Example 4.5], it was shown that a positive linear system with unbounded delays satisfying Assumption 3.2 may converge slower than any exponential function. However, an upper bound for the decay rate was not derived in [120]. Corollary 3.4 reveals that under Assumption 3.2 on delays, the decay rate of positive linear systems is upper bounded by a polynomial function of time.

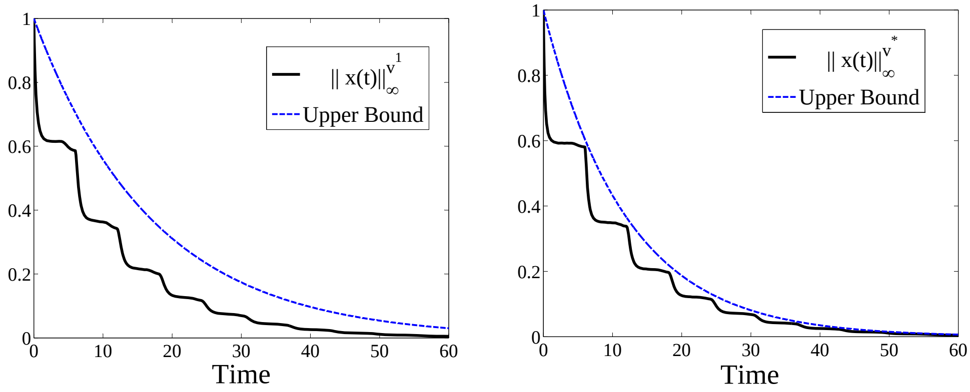


Figure 3.4: Comparison of upper bounds and actual decay rates of the positive linear system described by (3.27) without (left) and with (right) solving the convex optimization problem (3.26).

3.5 Discrete-time Homogeneous Order-preserving Systems

In this section, we study delay-independent stability of discrete-time positive systems whose vector fields are homogeneous and order-preserving.

3.5.1 Problem Statement

Consider the discrete-time analog of (3.1):

$$\Sigma : \begin{cases} x\big(k+1\big) = f\big(x(k)\big) + g\big(x(k-d(k))\big), & k \in \mathbb{N}_0, \\ \quad x\big(k\big) \quad = \varphi\big(k\big), & k \in \{-d_{\max}, \dots, 0\}. \end{cases} \tag{3.29}$$

Here, $x(k) \in \mathbb{R}^n$ is the system state, $f, g : \mathbb{R}^n \to \mathbb{R}^n$ are continuous vector fields with $f(\mathbf{0}) = g(\mathbf{0}) = \mathbf{0}$, $d_{\max} \in \mathbb{N}_0$, $\varphi : \{-d_{\max}, \dots, 0\} \to \mathbb{R}^n$ is the vector sequence specifying the initial state of the system, and $d(k)$ represents the time-varying delay which satisfies the following assumption.

Assumption 3.4. *The delay $d : \mathbb{N}_0 \to \mathbb{N}_0$ satisfies*

$$\lim_{k \to +\infty} k - d(k) = +\infty. \tag{3.30}$$

Intuitively, if Assumption 3.4 does not hold, computation of $x(k)$, even for large values of k, may involve the initial condition φ and those states near it, and hence $x(k)$ may not converge to zero as $k \to \infty$. To avoid this situation, Assumption (3.4) guarantees that old state information is eventually not used in evaluating (3.29). Note that Assumption (3.4) is the discrete-time counterpart of Assumption 3.1.

Remark 3.3. Assumption 3.4 implies that there exists a sufficiently large $T_0 \in \mathbb{N}_0$ such that $k - d(k) > 0$ for all $k > T_0$. Let

$$d_{\max} = -\inf_{0 \le k \le T_0} \left\{ k - d(k) \right\}.$$

Clearly, $d_{\max} \in \mathbb{N}_0$ is bounded. It follows that, even for unbounded delays satisfying Assumption 3.4, the initial condition φ is defined on a finite set $\{-d_{\max}, \ldots, 0\}$.

Definition 3.3. The time-delay system Σ given by (3.29) is said to be positive if for every nonnegative initial condition $\varphi(k) \ge 0$, $k \in \{-d_{\max}, \ldots, 0\}$, the corresponding solution is nonnegative, *i.e.*, $x(k) \ge \mathbf{0}$ for all $k \in \mathbb{N}$.

Positivity of discrete-time systems with time-varying delays is readily verified using the following result.

Proposition 3.2. *Consider the time-delay system Σ given by* (3.29). *If $f(x) \ge \mathbf{0}$ and $g(x) \ge \mathbf{0}$ for all $x \in \mathbb{R}^n_+$, then Σ is positive.*

Proof. See Appendix 3.7.6. ∎

For nonzero constant delays ($d(k) = d_{\max} > 0$, $k \in \mathbb{N}_0$), the sufficient condition in Proposition 3.2 is also necessary [58, Proposition 3.4]. However, the following example shows that this result may not true when delays are time-varying.

Example 3.7. Consider a discrete-time scalar system described by (3.29) with

$$f(x) = 2x, \ g(x) = -x, \ d(k) = \frac{1}{2}\left(1 - (-1)^k\right), \quad k \in \mathbb{N}_0.$$

Since $g(x) < 0$ for $x > 0$, the sufficient condition given in Proposition 3.2 is not satisfied. However, the solution of this system starting from the initial condition $x(0)$ is $x(k) = x(0)$, $k \in \mathbb{N}_0$, which implies that $x(k) \ge 0$ for all $x(0) \ge 0$. Therefore, this system with the time-varying delay $d(k)$ is positive. ∎

In the remainder of the section, f and g satisfy Assumption 3.5.

Assumption 3.5. *The following properties hold:*

1. *f and g are order-preserving on $\mathbb{R}^n_+$.*
2. *f and g are homogeneous of degree p with respect to the dilation map $\delta^r_\lambda(x)$.*

A system Σ given by (3.29) satisfying Assumption 3.5 is called *homogeneous order-preserving.* Since $f(\mathbf{0}) = g(\mathbf{0}) = \mathbf{0}$, Assumption 3.5.1 implies that f and g satisfy the positivity condition given in Proposition 3.2. Hence, homogeneous order-preserving systems are positive.

The main goal of this section is to study delay-independent stability of homogeneous order-preserving systems with time-varying delays satisfying Assumption 3.4.

3.5.2 Asymptotic Stability of Homogeneous Order-preserving Systems

The next theorem shows that under Assumption 3.4, the global asymptotic stability of order-preserving systems that are homogeneous of degree zero is insensitive to bounded and unbounded time-varying delays.

Theorem 3.3. *For the time-delay system* Σ *given by* (3.29)*, suppose that Assumptions 3.4 and 3.5 hold. Suppose also that* f *and* g *are homogeneous of degree zero. Then, the following statements are equivalent.*

(i) There exists a vector $v > \mathbf{0}$ *such that*

$$f(v) + g(v) < v. \tag{3.31}$$

(ii) Σ *is globally asymptotically stable for any nonnegative initial conditions and for all bounded and unbounded time-varying delays satisfying Assumption 3.4.*

(iii) Σ *without delay* ($d(k) = 0,\ k \in \mathbb{N}_0$) *is globally asymptotically stable for any nonnegative initial conditions.*

Proof. See Appendix 3.7.7. ∎

Theorem 3.3 provides a test for delay-independent stability of homogeneous order-preserving systems of degree zero; if we can demonstrate the existence of a vector $v > \mathbf{0}$ satisfying (3.31), then the origin is globally asymptotically stable for all delays satisfying Assumption 3.4. However, the following example illustrates that (3.31) is, in general, not a sufficient condition for global asymptotic stability of homogeneous order-preserving systems of degree greater than zero.

Example 3.8. Consider a discrete-time scalar system described by (3.29) with $f(x) = x^2$ and $g(x) = 0$. Clearly, f is order-preserving on $\mathbb{R}_+$ and homogeneous of degree one with respect to the standard dilation map. The solution of this system starting from the initial condition $x(0)$ is given by $x(k) = x(0)^{2^k}$, $k \in \mathbb{N}_0$. Thus, for initial conditions satisfying $x(0) \geq 1$, the system state $x(k)$ does not converge to the origin, *i.e.*, the origin is not globally asymptotically stable. However, since $f(0.5) = 0.25 < 0.5$, the stability condition (3.31) holds. ∎

We now show that under the stability condition (3.31), homogeneous order-preserving systems of degree greater than zero have a locally asymptotically stable equilibrium point at the origin, *i.e.*, $x(k)$ converges to the origin as $k \to \infty$ for sufficiently small initial conditions.

Corollary 3.5. *Consider the time-delay system* Σ *given by* (3.29)*. Suppose that Assumptions 3.4 and 3.5 hold and that* f *and* g *are homogeneous of degree greater than zero. If* (3.31) *admits a solution* $v > \mathbf{0}$*, then the origin is asymptotically stable with respect to initial conditions satisfying* $\mathbf{0} \leq \varphi(k) \leq v$*,* $k \in \{-d_{\max}, \ldots, 0\}$*.*

Proof. See Appendix 3.7.8. ∎

3.5.3 Decay Rates of Homogeneous Order-preserving Systems

The next definition introduces μ-stability for discrete-time systems.

Definition 3.4. *Suppose that $\mu : \mathbb{N} \to \mathbb{R}_+$ is a non-decreasing function satisfying $\mu(k) \to +\infty$ as $k \to +\infty$. The time-delay system Σ given by (3.29) is said to be globally μ-stable, if there exists a constant $M > 0$ such that for any initial condition φ, the solution $x(k)$ satisfies*

$$\|x(k)\| \leq \frac{M}{\mu(k)}, \quad k \in \mathbb{N},$$

where $\|\cdot\|$ is some norm on $\mathbb{R}^n$.

Paralleling our continuous-time results, global μ-stability of homogeneous order-preserving systems of degree zero can be established using the following theorem.

Theorem 3.4. *Consider the time-delay system Σ given by (3.29). Suppose that Assumptions 3.4 and 3.5 hold and that f and g are homogeneous of degree zero. In addition, assume that there is a function $\mu : \mathbb{N} \to \mathbb{R}_+$ such that the following conditions hold:*

(i) $\mu(k) > 0$, *for all* $k \in \mathbb{N}$.

(ii) $\mu(k+1) \geq \mu(k)$, *for all* $k \in \mathbb{N}$.

(iii) $\lim_{k\to+\infty} \mu(k) = +\infty$.

(iv) For each $i \in \{1, \ldots, n\}$,

$$\left(\lim_{k\to\infty} \frac{\mu(k+1)}{\mu(k)}\right)^{\frac{r_i}{r_{\max}}} \left(\frac{f_i(v)}{v_i}\right) + \left(\lim_{k\to\infty} \frac{\mu(k+1)}{\mu(k-d(k))}\right)^{\frac{r_i}{r_{\max}}} \left(\frac{g_i(v)}{v_i}\right) < 1,$$

where the vector $v > \mathbf{0}$ satisfies

$$f(v) + g(v) < v. \tag{3.32}$$

Then, every solution of Σ starting in the positive orthant satisfies

$$\left(\frac{x_i(k)}{v_i}\right)^{\frac{r_{\max}}{r_i}} = O\left(\mu^{-1}(k)\right), \quad k \in \mathbb{N},$$

for $i = 1, \ldots, n$.

Proof. See Appendix 3.7.9. ∎

Theorem 3.4 allows us to establish convergence rates of homogeneous order-preserving systems of degree zero under various classes of time-varying delays.

3.5.4 A Special Case: Discrete-time Positive Linear Systems

Let $f(x) = Ax$ and $g(x) = Bx$ such that $A, B \in \mathbb{R}^{n\times n}$ are nonnegative matrices. Then, the homogeneous order-preserving system (3.29) reduces to the positive linear system Σ_L of the form

$$\Sigma_L : \begin{cases} x\big(k+1\big) = Ax\big(k\big) + Bx\big(k - d(k)\big), & k \in \mathbb{N}_0, \\ \quad x\big(k\big) \quad = \varphi\big(k\big), & k \in \{-d_{\max}, \dots, 0\}. \end{cases} \tag{3.33}$$

Since Σ_L is homogeneous of degree zero, Theorem 3.3 can help us to derive a necessary and sufficient condition for delay-independent stability of positive linear systems. Specifically, we note the following.

Corollary 3.6. *Consider the time-delay linear system Σ_L given by* (3.33)*. Assume that A and B are nonnegative. Then, the positive linear system Σ_L is globally asymptotically stable for all time-delays satisfying Assumption 3.4 if and only if there exists a vector v such that*

$$\begin{cases} \big(A + B\big)v < v, \\ \qquad\quad v \;> \mathbf{0}. \end{cases} \tag{3.34}$$

Note that for the positive linear system (3.33), A and B are nonnegative, so $A + B$ is also nonnegative. According to Theorem 2.7, the linear programming problem (3.34) has a feasible solution v if and only if all eigenvalues of $A + B$ are strictly inside the unit circle.

3.6 Summary

This chapter has been concerned with delay-independent stability of a significant class of (continuous- and discrete-time) positive nonlinear systems with time-varying delays. We derived a set of necessary and sufficient conditions for global asymptotic stability of continuous-time homogeneous cooperative systems of arbitrary degree and discrete-time homogeneous order-preserving systems of degree zero with bounded and unbounded time-varying delays. These results show that the global asymptotic stability of such systems is independent of the magnitude and variation of time-delays. However, we also observed that the decay rates of these systems depend on how fast the delays can grow large. We developed two theorems for global μ-stability of homogeneous positive systems that quantify the convergence rates for various classes of time-delays. For positive linear systems, we further showed how the best convergence rates that our results guarantee can be found using convex optimization. For discrete-time homogeneous order-preserving systems of degree greater than zero, we demonstrated that the origin is locally asymptotically stable under the global asymptotic stability condition that we derived. Numerical examples justified the validity of our theoretical results.

3.7 Appendix

3.7.1 Proof of Proposition 3.1

Consider the following delayed differential equation

$$\begin{cases} \dot{y}(t) = f\big(y(t)\big) + g\big(y(t-\tau(t))\big) + \frac{1}{k}\mathbf{1}, & t \geq 0, \\ y(t) = \varphi(t), & t \in [-\tau_{\max}, 0], \end{cases} \tag{3.35}$$

where $k \in \mathbb{N}$, and $\mathbf{1} \in \mathbb{R}^n$ is the vector with all components equal to 1. Let $y^{(k)}(t)$ be the solution to (3.35) with the nonnegative initial condition $\varphi \in \mathcal{C}([-\tau_{\max}, 0], \mathbb{R}^n_+)$. Clearly, $y^{(k)}(0) = \varphi(0) \geq \mathbf{0}$. We claim that $y^{(k)}(t) \geq \mathbf{0}$ for all $t \geq 0$. By contradiction, suppose this is not true. Then, there exist an index $j \in \{1, \ldots, n\}$ and a time $t_1 \geq 0$ such that $y^{(k)}(t) \geq \mathbf{0}$ for all $t \in [0, t_1]$, $y_j^{(k)}(t_1) = 0$, and

$$D^+ y_j^{(k)}(t)\Big|_{t=t_1} \leq 0. \tag{3.36}$$

It follows from (3.5) and the above observations that

$$f_j\big(y^{(k)}(t_1)\big) \geq 0. \tag{3.37}$$

Since $t_1 - \tau(t_1) \in [-\tau_{\max}, t_1]$ and $\varphi(t) \geq \mathbf{0}$ for all $t \in [-\tau_{\max}, 0]$, we have

$$y^{(k)}\big(t_1 - \tau(t_1)\big) \geq \mathbf{0},$$

irrespectively of whether $t_1 - \tau(t_1)$ is nonnegative or not. From (3.5), we then obtain

$$g_j\big(y^{(k)}(t_1 - \tau(t_1))\big) \geq 0. \tag{3.38}$$

Using (3.37) and (3.38), the upper-right Dini-derivative of $y_j^{(k)}(t)$ along the trajectories of (3.35) at $t = t_1$ is given by

$$\begin{aligned} D^+ y_j^{(k)}(t)\Big|_{t=t_1} &= f_j\big(y^{(k)}(t_1)\big) + g_j\big(y^{(k)}(t_1 - \tau(t_1))\big) + \frac{1}{k} \\ &\geq \frac{1}{k} \\ &> 0, \end{aligned}$$

which contradicts (3.36). Therefore,

$$y^{(k)}(t) \geq \mathbf{0}, \quad \forall t \geq 0. \tag{3.39}$$

As k was an arbitrary natural number, it follows that (3.39) holds for all $k \in \mathbb{N}$. By letting $k \to \infty$, $y^{(k)}(t)$ converges to the solution $x(t)$ of (3.1) uniformly on $[-\tau_{\max}, \infty)$ [121, Theorem 2.1], which implies that $x(t) \geq \mathbf{0}$ for all $t \geq 0$. Hence, the time-delay system (3.1) is positive.

3.7.2 Proof of Theorem 3.1

$(i) \Rightarrow (ii)$: Let $v > \mathbf{0}$ be a vector such that (3.11) holds and let

$$\zeta = -\max_{1\le i\le n}\left\{\frac{f_i(v)+g_i(v)}{v_i}\right\}. \tag{3.40}$$

Since g is order-preserving on $\mathbb{R}^n_+$ and $g(\mathbf{0}) = \mathbf{0}$, $g_i(v) \ge 0$ for all i. Thus,

$$\frac{f_i(v)}{v_i} \le -\zeta, \tag{3.41}$$

for each $i = 1,\dots,n$. Define

$$\gamma_i = \left(1 + \frac{\zeta/2}{f_i(v)/v_i}\right)^{\frac{r_{\max}}{r_i+p}},$$

where $r_{\max} = \max_{1\le i\le n} r_i$. From (3.41), one can verify that $\gamma_i \in (0,1)$. We have

$$\begin{aligned}\gamma_i^{\frac{r_i+p}{r_{\max}}}\left(\frac{f_i(v)}{v_i}\right) + \frac{g_i(v)}{v_i} \quad &= \quad \frac{f_i(v)}{v_i} + \frac{g_i(v)}{v_i} + \frac{\zeta}{2}\\ &\le \quad -\frac{\zeta}{2},\end{aligned}$$

where we used (3.40) to get the inequality. For each i, it follows that

$$\gamma^{\frac{r_i+p}{r_{\max}}}\left(\frac{f_i(v)}{v_i}\right) + \frac{g_i(v)}{v_i} \le -\frac{\zeta}{2}, \tag{3.42}$$

where $\gamma = \max_{1\le i\le n}\gamma_i$. Clearly, $\gamma \in (0,1)$. The proof now proceeds in two steps:

1. First, we show that for any initial condition $\varphi \in \mathcal{C}([-\tau_{\max},0],\mathbb{R}^n_+)$, the corresponding solution $x(t)$ satisfies $x(t) \in S(0)$ for all $t \ge 0$, where the sets $S(m)$ are defined in (3.9).
2. By induction, we then prove that for each $m \in \mathbb{N}_0$, there exists $t_m \ge 0$ such that $x(t)$ will enter the set $S(m)$ at t_m and remains in this set for all $t \ge t_m$.

Step 1. Since the homogeneous cooperative system (3.1) is positive, $x_i(t) \ge 0$ for each $i = 1,\dots,n$ and all $t \ge 0$. Let

$$z_i(t) = \left(\frac{x_i(t)}{v_i}\right)^{\frac{r_{\max}}{r_i}} - \|\varphi\|, \tag{3.43}$$

where $\|\varphi\|$ is defined in (3.10). From the definition of $\|\varphi\|$, $z_i(0) \le 0$ for each i. We claim that $z_i(t) \le 0$ for all $t \ge 0$. By contradiction, suppose this is not true. Then, there exist an index $j \in \{1,\dots,n\}$ and a time $t_1 \ge 0$ such that

$$\begin{aligned} &z_i(t) \le 0, \quad i = 1,\dots,n,\ t \in [0,t_1],\\ &z_j(t_1) = 0, \end{aligned} \tag{3.44}$$

and

$$D^+ z_j(t)\Big|_{t=t_1} \geq 0. \tag{3.45}$$

From (3.43) and (3.44), we have

$$\begin{array}{rcl} x_i(t_1) & \leq & (\lambda_\varphi)^{r_i} v_i, \quad i = 1, \ldots, n, \; i \neq j, \\ x_j(t_1) & = & (\lambda_\varphi)^{r_j} v_j, \end{array}$$

where $\lambda_\varphi = \|\varphi\|^{\frac{1}{r_{\max}}}$. Since f is cooperative and homogeneous of degree p with respect to the dilation map $\delta_\lambda^r(x)$, it follows that

$$f_j\big(x(t_1)\big) \leq f_j\big(\delta^r_{\lambda_\varphi}(v)\big) = \big(\lambda_\varphi\big)^{r_j+p} f_j\big(v\big). \tag{3.46}$$

If $t_1 - \tau(t_1) \in [0, t_1]$, then, from (3.44), we have $z_i\big(t_1 - \tau(t_1)\big) \leq 0$, which implies that $x_i\big(t_1 - \tau(t_1)\big) \leq \big(\lambda_\varphi\big)^{r_i} v_i$ for each i, or, equivalently,

$$x\big(t_1 - \tau(t_1)\big) \leq \delta^r_{\lambda_\varphi}(v).$$

Note also that if $t_1 - \tau(t_1) \in [-\tau_{\max}, 0]$, then $x\big(t_1 - \tau(t_1)\big) = \varphi\big(t_1 - \tau(t_1)\big)$, and hence, from the definition of $\|\varphi\|$, the above inequality still holds. As g is order-preserving on $\mathbb{R}^n_+$ and homogeneous of degree p with respect to the dilation map $\delta^r_\lambda(x)$, this in turn implies that

$$g_j\big(x(t_1 - \tau(t_1))\big) \leq g_j\big(\delta^r_{\lambda_\varphi}(v)\big) = \big(\lambda_\varphi\big)^{r_j+p} g_j\big(v\big). \tag{3.47}$$

The upper-right Dini-derivative of $z_j(t)$ along the trajectories of (3.1) at $t = t_1$ is given by

$$\begin{array}{rcl} D^+ z_j(t)\Big|_{t=t_1} & = & \left(\dfrac{r_{\max}}{r_j}\right)\left(\dfrac{x_j(t_1)}{v_j}\right)^{(\frac{r_{\max}}{r_j}-1)} \left(\dfrac{f_j\big(x(t_1)\big) + g_j\big(x(t_1 - \tau(t_1))\big)}{v_j}\right) \\ & \leq & \left(\dfrac{r_{\max}}{r_j}\right)\Big(\lambda_\varphi\Big)^{r_{\max}-r_j}\Big(\lambda_\varphi\Big)^{r_j+p}\left(\dfrac{f_j(v) + g_j(v)}{v_j}\right) \\ & = & \left(\dfrac{r_{\max}}{r_j}\right)\Big(\lambda_\varphi\Big)^{r_{\max}+p}\left(\dfrac{f_j(v) + g_j(v)}{v_j}\right), \end{array}$$

where we used (3.46) and (3.47) to obtain the inequality. From (3.11), we conclude that $D^+ z_j(t_1) < 0$, which contradicts (3.45). Therefore, $z_i(t) \leq 0$ for all i and all $t \geq 0$, and hence $V(x(t)) \leq \|\varphi\|$ for $t \geq 0$. This shows that $x(t) \in S(0)$ for all $t \geq 0$.

Step 2. According to the previous step, the induction hypothesis is true for $m = 0$. Now, assume that it holds for a given m, *i.e.*, $V(x(t)) \leq \gamma^m \|\varphi\|$ for all $t \geq t_m$. We will prove that there exists a finite time $t_{m+1} \geq 0$ such that $x(t_{m+1}) \in S(m+1)$. By contradiction, suppose this is not true. Then,

$$\gamma^{m+1}\|\varphi\| \leq V(x(t)) \leq \gamma^m \|\varphi\|, \quad \forall t \geq t_m. \tag{3.48}$$

Let $i_t \in \{1, \ldots, n\}$ be an index such that $V(x(t)) = V_{i_t}(x_{i_t}(t))$, where

$$V_i(x_i) = \left(\frac{x_i}{v_i}\right)^{\frac{r_{\max}}{r_i}}.$$

The cooperativity and homogeneity of f implies that

$$\begin{aligned} f_{i_t}\big(x(t)\big) &\leq \big(V(x(t))\big)^{\left(\frac{r_{i_t}+p}{r_{\max}}\right)} f_{i_t}\big(v\big) \\ &\leq \big(\gamma^{m+1}\|\varphi\|\big)^{\left(\frac{r_{i_t}+p}{r_{\max}}\right)} f_{i_t}\big(v\big), \quad \forall t \geq t_m, \end{aligned} \tag{3.49}$$

where the second inequality follows from (3.48) and the fact that $f_{i_t}\big(v\big) < 0$. From Assumption 3.1, $\lim_{t\to\infty} t - \tau(t) = +\infty$. Thus, there exists sufficiently large $\hat{t}_m \geq t_m$ so that $t - \tau(t) \geq t_m$ for all $t \geq \hat{t}_m$. Since $x(t) \in S(m)$ for $t \geq t_m$, it follows that $x(t - \tau(t)) \in S(m)$ for all $t \geq \hat{t}_m$, implying that $V(x(t - \tau(t))) \leq \gamma^m\|\varphi\|$ for $t \geq \hat{t}_m$, or, equivalently,

$$x_i\big(t - \tau(t)\big) \leq \big(\gamma^m\|\varphi\|\big)^{\left(\frac{r_i}{r_{\max}}\right)} v_i, \quad \forall t \geq \hat{t}_m, \tag{3.50}$$

for all i. As g is order-preserving and homogeneous, we then have

$$g_{i_t}\big(x(t - \tau(t))\big) \leq \big(\gamma^m\|\varphi\|\big)^{\left(\frac{r_{i_t}+p}{r_{\max}}\right)} g_{i_t}\big(v\big), \quad \forall t \geq \hat{t}_m. \tag{3.51}$$

Substituting (3.49) and (3.51) into the upper-right Dini-derivative of $V_{i_t}(x_{i_t})$ along the trajectories of (3.1) yields

$$\begin{aligned} D^+V_{i_t}(x_{i_t}) &= \left(\frac{r_{\max}}{r_{i_t}}\right)\left(\frac{x_{i_t}(t)}{v_{i_t}}\right)^{(\frac{r_{\max}}{r_{i_t}}-1)}\left(\frac{f_{i_t}\big(x(t)\big) + g_{i_t}\big(x(t-\tau(t))\big)}{v_{i_t}}\right) \\ &\leq \left(\frac{r_{\max}}{r_{i_t}}\right)\left(\frac{x_{i_t}(t)}{v_{i_t}}\right)^{(\frac{r_{\max}}{r_{i_t}}-1)}\left(\gamma^m\|\varphi\|\right)^{\left(\frac{r_{i_t}+p}{r_{\max}}\right)}\left(\gamma^{\frac{r_{i_t}+p}{r_{\max}}}\left(\frac{f_{i_t}(v)}{v_{i_t}}\right) + \frac{g_{i_t}(v)}{v_{i_t}}\right) \\ &\leq -\left(\frac{r_{\max}}{r_{i_t}}\right)\left(\frac{x_{i_t}(t)}{v_{i_t}}\right)^{(\frac{r_{\max}}{r_{i_t}}-1)}\left(\gamma^m\|\varphi\|\right)^{\left(\frac{r_{i_t}+p}{r_{\max}}\right)}\left(\frac{\zeta}{2}\right), \\ &\leq \underbrace{-\left(\frac{r_{\max}}{r_{i_t}}\right)\left(\gamma^{m+1}\|\varphi\|\right)^{(1-\frac{r_{i_t}}{r_{\max}})}\left(\gamma^m\|\varphi\|\right)^{\left(\frac{r_{i_t}+p}{r_{\max}}\right)}\left(\frac{\zeta}{2}\right)}_{\kappa}, \quad \forall t \geq \hat{t}_m \end{aligned} \tag{3.52}$$

where the last two inequalities follow from (3.42) and (3.48), respectively. Note that $\kappa > 0$. Since $V_i(x_i)$ is continuously differentiable on $\mathbb{R}$ for each i, $V(x)$ is locally Lipschitz and

$$D^+V(x(t)) = \max_{j\in\mathcal{J}(t)} D^+V_j(x_j(t)),$$

where $\mathcal{J}(t) = \big\{i \mid V_i(x_i(t)) = V(x(t))\big\}$ [122]. It follows from (3.52) that

$$D^+V(x(t)) \leq -\kappa, \quad \forall t \geq \hat{t}_m.$$

This together with (3.48) implies that

$$\begin{aligned} V(x(t)) &\leq V(x(\hat{t}_m)) - \kappa(t - \hat{t}_m) \\ &\leq \gamma^m \|\varphi\| - \kappa(t - \hat{t}_m), \quad \forall t \geq \hat{t}_m. \end{aligned}$$

It is immediate to see that the the right-hand side of the above inequality becomes smaller than $\gamma^{m+1}\|\varphi\|$ when

$$t \geq t_{m+1} = \hat{t}_m + \gamma^m \|\varphi\| \frac{1-\gamma}{\kappa},$$

which contradicts (3.48). Thus necessarily, $x(t)$ reaches $S(m+1)$ in a finite time.

We now prove that $x(t)$ remains in $S(m+1)$ for all $t \geq t_{m+1}$. Let

$$w_i(t) = \left(\frac{x_i(t)}{v_i}\right)^{\frac{r_{\max}}{r_i}} - \gamma^{m+1}\|\varphi\|, \quad t \geq t_{m+1}. \tag{3.53}$$

Since $x(t_{m+1}) \in S(m+1)$, $w_i(t_{m+1}) \leq 0$ for all i. We show that $w_i(t) \leq 0$ for all i and all $t \geq t_{m+1}$. If, by contradiction, this is not true, then there is an index $j \in \{1, \ldots, n\}$ and a time $t_2 \geq t_{m+1}$ such that $w_i(t) \leq 0$ for $t \in [t_{m+1}, t_2]$, $w_j(t_2) = 0$, and

$$D^+ w_j(t)\Big|_{t=t_2} \geq 0. \tag{3.54}$$

From (3.53), we have

$$\begin{aligned} x_i(t_2) &\leq \left(\gamma^{m+1}\|\varphi\|\right)^{\frac{r_i}{r_{\max}}} v_i, \quad i = 1, \ldots, n,\ i \neq j, \\ x_j(t_2) &= \left(\gamma^{m+1}\|\varphi\|\right)^{\frac{r_j}{r_{\max}}} v_j. \end{aligned}$$

It now follows from cooperativity and homogeneity of f that

$$f_j\big(x(t_2)\big) \leq \left(\gamma^{m+1}\|\varphi\|\right)^{\frac{r_j+p}{r_{\max}}} f_j\big(v\big). \tag{3.55}$$

Moreover, since $t_2 \geq t_{m+1} \geq \hat{t}_m$, it follows from (3.50) that

$$g_j\big(x(t_2 - \tau(t_2))\big) \leq \left(\gamma^m \|\varphi\|\right)^{\frac{r_j+p}{r_{\max}}} g_j\big(v\big), \tag{3.56}$$

where we used the fact that g is order-preserving and homogeneous. The upper-right Dini-derivative of $w_j(t)$ along the trajectories of (3.1) at $t = t_2$ is given by

$$\begin{aligned} D^+ w_j(t)\Big|_{t=t_2} &\leq \left(\frac{r_{\max}}{r_j}\right)\left(\frac{x_j(t_2)}{v_j}\right)^{\left(\frac{r_{\max}}{r_j}-1\right)} \left(\gamma^m\|\varphi\|\right)^{\frac{r_j+p}{r_{\max}}} \left(\gamma^{\frac{r_j+p}{r_{\max}}}\left(\frac{f_j(v)}{v_j}\right) + \frac{g_j(v)}{v_j}\right) \\ &< 0, \end{aligned}$$

where we used (3.55) and (3.56) to get the first inequality and (3.42) to obtain the second inequality. This contradicts (3.54), and hence $w_i(t) \leq 0$ for all i and all $t \geq t_{m+1}$. It follows that $V(x(t)) \leq \gamma^{m+1}\|\varphi\|$ for $t \geq t_{m+1}$, or, equivalently, $x(t) \in S(m+1)$ for all $t \geq t_{m+1}$.

In summary, we conclude that for each $m \in \mathbb{N}_0$, there exists $t_m \geq 0$ such that $x(t) \in S(m)$ for all $t \geq t_m$. Since $\gamma \in (0,1)$, γ^m approaches zero as $m \to \infty$. Therefore, the origin is globally asymptotically stable.

$(ii) \Rightarrow (iii)$: Assume that the time-delay system (3.1) is globally asymptotically stable for all delays satisfying Assumption 3.1. Particularly, let $\tau(t) = 0$. Then, the non-delayed system $\dot{x}(t) = f(x(t)) + g(x(t))$ is globally asymptotically stable.

$(iii) \Rightarrow (i)$: As $f+g$ is a cooperative vector field, it follows from [123, Proposition 3.10, Theorem 3.12] that there is a vector $v > \mathbf{0}$ such that (3.11) holds.

3.7.3 Proof of Theorem 3.2

Let $v > \mathbf{0}$ be a vector satisfying (3.13). According to Theorem 3.1, the homogeneous cooperative system (3.1) is globally asymptotically stable for all nonnegative initial conditions and for all delays satisfying Assumption 3.1. We will prove that (3.1) is also globally μ-stable. From Remark 3.1, there exists $T_0 > 0$ large enough such that

$$t - \tau(t) > 0, \quad \forall t > T_0. \tag{3.57}$$

By condition (iv), we can find a sufficiently large constant $T_1 > 0$ such that for all $t > T_1$ and all $i \in \{1, \ldots, n\}$,

$$\left(\frac{r_{\max}}{r_i}\right)\left(\left(\frac{f_i(v)}{v_i}\right) + \left(\frac{\mu(t)}{\mu(t-\tau(t))}\right)^{\frac{r_i+p}{r_{\max}}}\left(\frac{g_i(v)}{v_i}\right)\right) + \frac{\dot{\mu}(t)}{(\mu(t))^{1-\frac{p}{r_{\max}}}} < 0.$$

Since $\mu(t)$ is non-decreasing on $\mathbb{R}_+$, it follows that

$$\varepsilon\left(\frac{r_{\max}}{r_i}\right)\left(\left(\frac{f_i(v)}{v_i}\right) + \left(\frac{\mu(t)}{\mu(t-\tau(t))}\right)^{\frac{r_i+p}{r_{\max}}}\left(\frac{g_i(v)}{v_i}\right)\right) + \frac{\dot{\mu}(t)}{(\mu(t))^{1-\frac{p}{r_{\max}}}} < 0, \tag{3.58}$$

holds for any $\varepsilon \geq 1$. Let $M = \max\{1, \mu(\overline{T})\|\varphi\|\}$, where $\overline{T} = \max\{T_0, T_1\} + 1$, and $\|\varphi\|$ is defined in (3.10). According to the proof of Theorem 3.1, $V(x(t)) \leq \|\varphi\|$ for all $t \geq 0$. Thus,

$$\begin{aligned}\sup_{0\leq t\leq \overline{T}}\{\mu(t)V(x(t))\} &\leq \sup_{0\leq t\leq \overline{T}}\{\mu(t)\}\|\varphi\| \\ &= \mu(\overline{T})\|\varphi\| \\ &\leq M, \end{aligned}\tag{3.59}$$

where we used condition (ii) to get the equality. It follows that

$$\mu(t)V(x(t)) \leq M, \quad \forall t \in [0, \overline{T}]. \tag{3.60}$$

In order to prove global μ-stability, we will show that (3.60) also holds for all $t \geq \overline{T}$. By contradiction, suppose this is not true. Then, there exist an index $j \in \{1, \ldots, n\}$ and a time $t_1 \geq \overline{T}$ such that

$$\mu(t)V(x(t)) \leq M, \quad t \in [0, t_1], \tag{3.61}$$

$$\mu(t_1)\left(\frac{x_j(t_1)}{v_j}\right)^{\frac{r_{\max}}{r_j}} = M, \tag{3.62}$$

$$D^+\mu(t)\left(\frac{x_j(t)}{v_j}\right)^{\frac{r_{\max}}{r_j}}\Bigg|_{t=t_1} \geq 0. \tag{3.63}$$

From (3.61) and (3.62), we have

$$\begin{aligned} x_i(t_1) &\leq \left(\frac{M}{\mu(t_1)}\right)^{\frac{r_i}{r_{\max}}} v_i, \quad i = 1, \ldots, n,\ i \neq j, \\ x_j(t_1) &= \left(\frac{M}{\mu(t_1)}\right)^{\frac{r_j}{r_{\max}}} v_j. \end{aligned}$$

As f is cooperative and homogeneous, it follows that

$$f_j\big(x(t_1)\big) \leq \left(\frac{M}{\mu(t_1)}\right)^{\frac{r_j+p}{r_{\max}}} f_j(v). \tag{3.64}$$

From (3.57), since $t_1 \geq \overline{T} > T_0$, we have $t_1 \geq t_1 - \tau(t_1) > 0$. Hence, from (3.61),

$$\mu\big(t_1 - \tau(t_1)\big)V(x(t_1 - \tau(t_1))) \leq M.$$

As g is order-preserving and homogeneous, this in turn implies

$$g_j\big(x(t_1 - \tau(t_1))\big) \leq \left(\frac{M}{\mu(t_1 - \tau(t_1))}\right)^{\frac{r_j+p}{r_{\max}}} g_j(v). \tag{3.65}$$

We then have

$$\begin{aligned} & D^+\mu(t)\left(\frac{x_j(t)}{v_j}\right)^{\frac{r_{\max}}{r_j}}\Bigg|_{t=t_1} \\ &= \mu(t_1)\left(\frac{r_{\max}}{r_j}\right)\left(\frac{x_j(t_1)}{v_j}\right)^{(\frac{r_{\max}}{r_j}-1)}\frac{\dot{x}_j(t_1)}{v_j} + \dot{\mu}(t_1)\left(\frac{x_j(t_1)}{v_j}\right)^{\frac{r_{\max}}{r_j}} \\ &= \mu(t_1)\left(\frac{r_{\max}}{r_j}\right)\left(\frac{M}{\mu(t_1)}\right)^{(1-\frac{r_j}{r_{\max}})}\left(\frac{f_j(x(t_1)) + g_j(x(t_1 - \tau(t_1)))}{v_j}\right) + M\frac{\dot{\mu}(t_1)}{\mu(t_1)} \\ &\leq \frac{M}{(u(t_1))^{\frac{p}{r_{\max}}}} \times \\ &\quad \left\{M^{\frac{p}{r_{\max}}}\left(\frac{r_{\max}}{r_j}\right)\left(\left(\frac{f_j(v)}{v_j}\right) + \left(\frac{\mu(t_1)}{\mu(t_1 - \tau(t_1))}\right)^{\frac{r_j+p}{r_{\max}}}\left(\frac{g_j(v)}{v_j}\right)\right) + \frac{\dot{\mu}(t_1)}{\mu^{1-\frac{p}{r_{\max}}}}\right\}, \end{aligned}$$

where we used (3.62) to get the second equality, and (3.64)–(3.65) to obtain the inequality. Since $M \geq 1$ and $t_1 \geq \overline{T} > T_1$, it now follows from (3.58) that

$$D^+\mu(t)\left(\frac{x_j(t)}{v_j}\right)^{\frac{r_{\max}}{r_j}}\Bigg|_{t=t_1} < 0,$$

which contradicts (3.63). We conclude that $\mu(t)V(x(t)) \leq M$ for all $t \geq \overline{T}$, and hence

$$V(x(t)) \leq \frac{M}{\mu(t)}, \quad t \geq 0.$$

This completes the proof.

3.7.4 Proof of Corollary 3.1

(i) Assume that $p = 0$. For each $i = 1, \ldots, n$, equation (3.16) has a unique positive solution η_i. Pick a constant η satisfying $\eta \in (0, \min_{1\leq i\leq n} \eta_i)$. Since the left-hand side of (3.16) is strictly monotonically increasing in $\eta_i > 0$, we have

$$\left(\frac{r_{\max}}{r_i}\right)\left(\left(\frac{f_i(v)}{v_i}\right) + \left(e^{\eta\tau_{\max}}\right)^{\frac{r_i}{r_{\max}}}\left(\frac{g_i(v)}{v_i}\right)\right) + \eta < 0, \quad i = 1, \ldots, n. \tag{3.66}$$

Now, let $\mu(t) = e^{\eta t}$. One can verify that $\mu(t)$ satisfies conditions (i)–(iii) of Theorem 3.2. Moreover,

$$\lim_{t\to\infty}\frac{\dot{\mu}(t)}{\mu(t)} = \eta,$$

and

$$\lim_{t\to\infty}\frac{\mu(t)}{\mu(t-\tau(t))} \leq \lim_{t\to\infty}\frac{e^{\eta t}}{e^{\eta(t-\tau_{\max})}} = e^{\eta\tau_{\max}},$$

where the inequality holds since $\tau(t) \leq \tau_{\max}$ and $\mu(t)$ is non-decreasing. It follows from (3.66) and the above observations that condition (iv) of Theorem 3.2 is also satisfied. Hence, the solution $x(t)$ of (3.1) satisfies (3.15).

(ii) If $p > 0$, we can pick $\mu(t) = (\theta t + 1)^{\frac{r_{\max}}{p}}$. The rest of the proof is similar to the one for $p = 0$ and thus omitted.

3.7.5 Proof of Corollary 3.2

(i) Assume that $p = 0$. The left-hand side of (3.20) is strictly monotonically increasing in $\xi_i > 0$. Thus,

$$\left(\frac{f_i(v)}{v_i}\right) + \left(\frac{1}{1-\alpha}\right)^{\frac{r_i}{r_{\max}}\xi}\left(\frac{g_i(v)}{v_i}\right) < 0, \quad i = 1, \ldots, n,$$

where $\xi \in \left(0, \min_{1\leq i\leq n}\xi_i\right)$. Now, letting $\mu(t) = t^{\xi}$, the rest of the proof is similar to the one of Corollary 3.1 and thus omitted.

(ii) When $p > 0$, we can choose $\mu(t) = t^{\frac{r_{\max}}{p}\beta}$, with $\beta \in (0,1)$ satisfying (3.21).

3.7.6 Proof of Proposition 3.2

Let φ be a nonnegative initial condition, *i.e.*, $\varphi(k) \geq \mathbf{0}$ for $k \in \{-d_{\max}, \ldots, 0\}$. We show by induction that $x(k) \geq \mathbf{0}$ for all $k \in \mathbb{N}_0$. Since $x(0) = \varphi(0)$, the induction hypothesis is true for $k = 0$. Assume for induction that $x(k) \geq \mathbf{0}$ for $k \in \{0, \ldots, \overline{k}\}$ with $\overline{k} \in \mathbb{N}_0$. Clearly, $x(\overline{k}) \geq \mathbf{0}$. Moreover, since $\overline{k} - d(\overline{k}) \in [-d_{\max}, \overline{k}]$ and $\varphi(k) \geq \mathbf{0}$, we have $x(\overline{k} - d(\overline{k})) \geq \mathbf{0}$. As $f(x) \geq \mathbf{0}$ and $g(x) \geq \mathbf{0}$ for all $x \in \mathbb{R}^n_+$, it follows that $f\big(x(\overline{k})\big) \geq \mathbf{0}$ and $g\big(x(\overline{k} - d(\overline{k}))\big) \geq \mathbf{0}$. Therefore,

$$x\big(\overline{k} + 1\big) = f\big(x(\overline{k})\big) + g\big(x(\overline{k} - d(\overline{k}))\big) \geq \mathbf{0}.$$

By induction, we conclude that $x(k) \geq \mathbf{0}$ for all $k \in \mathbb{N}_0$. Hence, (3.29) is positive.

3.7.7 Proof of Theorem 3.3

$(i) \Rightarrow (ii)$: Let $v > \mathbf{0}$ be a vector such that (3.31) holds and let

$$\gamma = \max_{1 \leq i \leq n} \left(\frac{f_i(v) + g_i(v)}{v_i} \right)^{\frac{r_{\max}}{r_i}}. \tag{3.67}$$

Note that $\gamma \in (0, 1)$. Consider the homogeneous order-preserving system (3.29) without time-delays, given by

$$x\big(k + 1\big) = f\big(x(k)\big) + g\big(x(k)\big), \quad k \in \mathbb{N}_0. \tag{3.68}$$

First, we use induction to show that for the initial condition $x(0) = v$, the solution $x(k)$ to (3.68) satisfies

$$\left(\frac{x_i(k)}{v_i} \right)^{\frac{r_{\max}}{r_i}} \leq \gamma^k, \quad k \in \mathbb{N}_0,$$

for each $i = 1, \ldots, n$. The induction hypothesis is true for $k = 0$, since $x_i(0) = v_i$ for each i. Assuming it is true for a given $k = \overline{k} \in \mathbb{N}_0$. Then,

$$x_i(\overline{k}) \leq \gamma^{\frac{r_i \overline{k}}{r_{\max}}} v_i, \quad i = 1, \ldots, n.$$

As f and g are homogeneous of degree zero and order-preserving, it follows that

$$\begin{aligned} \frac{x_i(\overline{k} + 1)}{v_i} &= \frac{f_i\big(x(\overline{k})\big) + g_i\big(x(\overline{k})\big)}{v_i} \\ &\leq \gamma^{\frac{r_i \overline{k}}{r_{\max}}} \left(\frac{f_i(v) + g_i(v)}{v_i} \right) \\ &\leq \gamma^{\frac{r_i (\overline{k}+1)}{r_{\max}}}, \end{aligned}$$

where we used (3.67) to obtain the second inequality. Therefore,

$$\left(\frac{x_i(\overline{k}+1)}{v_i}\right)^{\frac{r_{\max}}{r_i}} \leq \gamma^{\overline{k}+1},$$

which completes the induction proof. Since $\gamma \in (0,1)$, γ^k approaches zero as $k \to +\infty$. Hence, for the initial condition $x(0) = v$, the solution $x(k)$ to the non-delayed system (3.68) converges asymptotically to the origin. We now prove that the homogeneous order-preserving system (3.29) with time-delays is globally asymptotically stable. Since

$$\mathbf{0} \leq f(\mathbf{0}) + g(\mathbf{0}) \leq f(v) + g(v) \leq v,$$

it follows from the asynchronous convergence theorem for totally asynchronous iterations [16, §6.4] that the solution $x(k)$ to the time-delay system (3.29) starting from initial conditions $\varphi(k) \in [\mathbf{0}, v]$ converges asymptotically to the origin. Let φ be an arbitrary nonnegative initial condition (φ may not satisfy $\varphi(k) \in [\mathbf{0}, v]$). As $v > \mathbf{0}$, there exists $\lambda \geq 1$ sufficiently large such that

$$\varphi(k) \leq \delta_\lambda^r(v) = (\lambda^{r_1} v_1, \ldots, \lambda^{r_n} v_n),$$

for all $k \in \{-d_{\max}, \ldots, 0\}$. Let $\overline{v} = \delta_\gamma^r(v)$. Since f and g are homogeneous of degree zero, we have

$$\begin{aligned} f_i(\overline{v}) + g_i(\overline{v}) &= \lambda^{r_i}\left(f_i(v) + g_i(v)\right) \\ &< \lambda^{r_i} v_i \\ &= \overline{v}_i, \end{aligned}$$

where we used (3.31) to get the inequality. Thus, $f(\overline{v}) + g(\overline{v}) < \overline{v}$. We conclude that for any nonnegative initial condition φ, there exist a vector $\overline{v} > \mathbf{0}$ such that (3.31) holds and that $\varphi(k) \in [\mathbf{0}, \overline{v}]$, $k \in \{-d_{\max}, \ldots, 0\}$. It follows that the homogeneous order-preserving system (3.29) is globally asymptotically stable.

$(ii) \Rightarrow (iii)$: Suppose that (3.29) is globally asymptotically stable for all delays satisfying Assumption 3.4. Particularly, let $d(k) = 0$. Then, the non-delayed system $x\big(k+1\big) = f\big(x(k)\big) + g\big(x(k)\big)$ is globally asymptotically stable.

$(iii) \Rightarrow (i)$: Since $f + g$ is continuous, order-preserving on $\mathbb{R}_+^n$ and $(f+g)(\mathbf{0}) = \mathbf{0}$, the conclusion follows from [124, Propositions 5.2 and 5.6].

3.7.8 Proof of Corollary 3.5

The proof is similar to the one of Theorem 3.3 and thus omitted.

3.7.9 Proof of Theorem 3.4

Let $v > \mathbf{0}$ be a vector satisfying (3.32). According to Theorem 3.3, the homogeneous order-preserving system (3.29) with time-delays satisfying Assumption 3.4 is globally asymptotically stable. We will prove that (3.29) is also globally μ-stable. From Remark 3.3, there exists $T_0 \in \mathbb{N}$ such that

$$k - d(k) > 0, \quad \forall k > T_0. \tag{3.69}$$

From condition (iv), one can find a sufficiently large constant $T_1 \in \mathbb{N}$, such that for all $k > T_1$, we have

$$\left(\frac{\mu(k+1)}{\mu(k)}\right)^{\frac{r_i}{r_{\max}}} \left(\frac{f_i(v)}{v_i}\right) + \left(\frac{\mu(k+1)}{\mu(k-d(k))}\right)^{\frac{r_i}{r_{\max}}} \left(\frac{g_i(v)}{v_i}\right) < 1. \tag{3.70}$$

Let $M = \mu(\overline{T})\|\varphi\|$, where $\overline{T} = \max\{T_0, T_1\} + 1$, and $\|\varphi\|$ is defined in (3.10). We now use induction to prove that

$$V(x(k)) \leq \frac{M}{\mu(k)}, \quad \forall k \in \mathbb{N}. \tag{3.71}$$

According to the proof of Theorem 3.3, $V(x(k)) \leq \|\varphi\|$ for all $k \in \mathbb{N}$. Thus,

$$\begin{aligned} \max_{1\leq k\leq \overline{T}} \{\mu(k)V(x(k))\} &\leq \max_{1\leq k\leq \overline{T}} \{\mu(k)\}\|\varphi\| \\ &= \mu(\overline{T})\|\varphi\| \\ &= M, \end{aligned} \tag{3.72}$$

where we used condition (ii) to get the first equality. It follows from (3.72) that (3.71) is true for $k \in \{1, \ldots, \overline{T}\}$. Next, assume for induction that (3.71) holds for all k up to some $\overline{k}$, where $\overline{k} \geq \overline{T}$. Thus,

$$0 \leq \left(\frac{x_i(k)}{v_i}\right)^{\frac{r_{\max}}{r_i}} \leq \frac{M}{\mu(k)}, \quad k = 1, \ldots, \overline{k},$$

which implies that

$$0 \leq x_i(\overline{k}) \leq \left(\frac{M}{\mu(\overline{k})}\right)^{\frac{r_i}{r_{\max}}} v_i. \tag{3.73}$$

Since $\overline{k} \geq \overline{T} > T_0$, it follows from (3.69) that $\overline{k} - d(\overline{k}) \in \{1, \ldots, \overline{k}\}$. Hence,

$$0 \leq x_i\big(\overline{k} - d(\overline{k})\big) \leq \left(\frac{M}{\mu(\overline{k} - d(\overline{k}))}\right)^{\frac{r_i}{r_{\max}}} v_i. \tag{3.74}$$

As f and g are homogeneous of degree zero and order-preserving on $\mathbb{R}^n_+$, it follows from (3.73) and (3.74) that

$$\begin{aligned} f_i\big(x(\overline{k})\big) &\leq \left(\frac{M}{\mu(\overline{k})}\right)^{\frac{r_i}{r_{\max}}} f_i\big(v\big), \\ g_i\big(x(\overline{k} - d(\overline{k})\big) &\leq \left(\frac{M}{\mu(\overline{k} - d(\overline{k}))}\right)^{\frac{r_i}{r_{\max}}} g_i\big(v\big). \end{aligned} \tag{3.75}$$

We now show that $x(\overline{k}+1)$ satisfies (3.71). For each $i \in \{1,\dots,n\}$, we have

$$\begin{aligned}
\frac{x_i(\overline{k}+1)}{v_i} &= \frac{f_i\big(x(\overline{k})\big) + g_i\big(x(\overline{k}-d(\overline{k}))\big)}{v_i} \\
&\le \left(\frac{M}{\mu(\overline{k})}\right)^{\frac{r_i}{r_{\max}}} \left(\frac{f_i(v)}{v_i}\right) + \left(\frac{M}{\mu(\overline{k}-d(\overline{k}))}\right)^{\frac{r_i}{r_{\max}}} \left(\frac{g_i(v)}{v_i}\right) \\
&\le \left(\frac{M}{\mu(\overline{k}+1)}\right)^{\frac{r_i}{r_{\max}}},
\end{aligned}$$

where we used (3.75) to get the first inequality and (3.70) to obtain the second inequality. Therefore,

$$V(x(\overline{k}+1)) \le \frac{M}{\mu(\overline{k}+1)},$$

and hence the induction proof is complete.

Chapter 4

Monotone Systems with Heterogonous Delays

SUB-HOMOGENEOUS positive monotone systems constitute an important and useful class of positive nonlinear systems. Established models of many physical phenomena fall within this class. For example, biochemical reaction networks and most power control algorithms in wireless networks can be analyzed as sub-homogeneous positive monotone systems [125–127]. This class of positive systems includes homogeneous positive systems as a special case.

Many processes that are described by positive systems are comprised of interconnected subsystems that exchange material, energy, or information. As components could operate asynchronously and transfer between subsystems typically takes time, accurate models of such systems must include time-delays. Recently, it has been shown that the asymptotic stability of sub-homogeneous positive monotone systems is independent of *constant* delays [128]. In practice, however, delays are often *time-varying*. Hence, a natural question is if sub-homogeneous positive monotone systems are insensitive also to time-varying delays. In the previous chapter, we have shown that the asymptotic stability of homogeneous positive systems is independent of the magnitude and variation of time-varying delays. Therefore, it is reasonable to conjecture that sub-homogeneous positive systems are insensitive to time-varying delays, at least as long as the delays are bounded. Proving or disproving the conjecture is nontrivial. The main reason for this is that the homogeneity assumption, which played a key role in our stability analysis of homogeneous positive systems with time-varying delays, is not satisfied for sub-homogeneous positive systems.

Contributions of the Chapter. In this chapter, we show that the conjecture is true. By transforming the stability problem with *heterogeneous time-varying* delays into one with constant delays, we demonstrate that a sub-homogeneous positive monotone system with arbitrary bounded heterogeneous time-varying delays is globally asymptotically stable if and only if the corresponding system without delay is globally asymptotically stable. More specifically, we make the following contributions. First, we derive a sufficient condition for asymptotic stability of general monotone (not necessarily sub-homogeneous) systems with time-varying delays. The proof technique is based on an extension of a delay-independent stability result for monotone systems under constant delays by Smith [69] to systems with bounded heterogenous time-varying delays. Under the additional assumption of positivity and sub-homogeneity of vector fields, we establish the aforementioned delay insensitivity property and derive a novel test for global asymptotic stability. If a sub-homogeneous positive system has a unique equilibrium point in the positive orthant, we prove that our stability test is necessary and sufficient. Specialized to positive linear systems, our results extend and sharpen existing results from the literature. Since sub-homogeneous positive systems include homogeneous positive systems as a special case, our work also generalizes the asymptotic stability results of the previous chapter for the case of bounded time-varying delays.

Outline of the Chapter. Section 4.1 formulates the problem that we address in this chapter. Section 4.2 presents our main results on the delay-independent stability of monotone systems and the insensitivity of sub-homogeneous positive monotone systems to bounded heterogeneous time-varying delays. Illustrative examples are included throughout the development of the results. Finally, conclusions are given in Section 4.3. In the appendix, we provide detailed proofs of the main results.

4.1 Problem Statement

We consider the following nonlinear dynamical system with heterogeneous time-varying delays

$$\mathcal{G}: \quad \begin{cases} \dot{x}_i\big(t\big) = f_i\big(x(t)\big) + g_i\big(x_1(t-\tau_1^i(t)),\dots,x_n(t-\tau_n^i(t))\big), & t \geq 0, \\ x_i\big(t\big) = \varphi_i\big(t\big), & t \in [-\tau_{\max},0]. \end{cases} \tag{4.1}$$

Here, $i = 1,\dots,n$, $x(t) = (x_1(t),\dots,x_n(t)) \in \mathbb{R}^n$ is the state vector, $\tau_{\max} \in \mathbb{R}_+$, $f(x) = (f_1(x),\dots,f_n(x))$ and $g(x) = (g_1(x),\dots,g_n(x))$ are continuously differentiable vector fields on the convex set $\mathcal{W} \subseteq \mathbb{R}^n$, and

$$\varphi(t) = (\varphi_1(t),\dots,\varphi_n(t)) \in \mathcal{C}\big([-\tau_{\max},0],\mathcal{W}\big),$$

is the vector-valued function specifying the initial condition of the system. The time-varying delays $\tau_j^i(t)$ are continuous with respect to time, and satisfy

$$0 \leq \tau_j^i(t) \leq \tau_{\max}, \quad t \geq 0,$$

for all $i,j = 1,\dots,n$. Note that the maximum delay bound $\tau_{\max}$ may be unknown, that $\tau_j^i(t)$ are not necessarily continuously differentiable, and that no restriction on their derivative is imposed. Since the initial condition φ and the delays $\tau_j^i(t)$ are continuous functions of time, the existence and uniqueness of solutions to (4.1) follow from [99, Theorem 2.3]. We denote the unique solution of (4.1) corresponding to the initial condition $\varphi(t)$, $t \in [-\tau_{\max},0]$, by $x(t,\varphi)$. The equilibria of the time-delay system $\mathcal{G}$ given by (4.1) are defined as constant functions $\varphi(t) = x^\star$, $t \in [-\tau_{\max}, 0]$, where the vector $x^\star \in \mathcal{W}$ satisfies

$$f(x^\star) + g(x^\star) = \mathbf{0}. \tag{4.2}$$

In general, (4.2) may have more than one solution $x^\star$ and, hence, $\mathcal{G}$ may have multiple equilibrium points.

In this chapter, we study the asymptotic stability of time-delay systems of the form (4.1) which are *monotone*:

Definition 4.1. The time-delay system $\mathcal{G}$ given by (4.1) is called monotone if for any initial conditions $\varphi, \varphi' \in \mathcal{C}\big([-\tau_{\max}, 0], \mathcal{W}\big)$ satisfying

$$\varphi(t) \leq \varphi'(t), \quad t \in [-\tau_{\max}, 0],$$

we have $x(t, \varphi) \leq x(t, \varphi')$ for all $t \geq 0$.

Loosely speaking, the trajectories of a monotone system starting at ordered initial conditions preserve the same ordering during the time evolution. Monotonicity of time-delay systems is readily verified using the next result.

Proposition 4.1 (**[69, Theorem 5.1.1]**). *Consider the time-delay system $\mathcal{G}$ given by* (4.1). *Suppose that f is cooperative and g is order-preserving on $\mathcal{W}$. Then, $\mathcal{G}$ is monotone.*

We now provide a necessary and sufficient condition for positivity of monotone systems with heterogeneous time-varying delays. Recall that the time-delay system $\mathcal{G}$ given by (4.1) is said to be positive if for any nonnegative initial condition $\varphi \in \mathcal{C}\big([-\tau_{\max}, 0], \mathbb{R}^n_+\big)$, the corresponding state trajectory will remain in the positive orthant, *i.e.*, $x(t, \varphi) \in \mathbb{R}^n_+$ for all $t \geq 0$.

Proposition 4.2. *For the time-delay system $\mathcal{G}$ given by* (4.1)*, assume that f is cooperative on $\mathbb{R}^n_+$ and g is order-preserving on $\mathbb{R}^n_+$. Then, the monotone system $\mathcal{G}$ is positive if and only if*

$$f(\mathbf{0}) + g(\mathbf{0}) \geq \mathbf{0}. \tag{4.3}$$

Proof. See Appendix 4.4.2. ■

Proposition 4.2 shows that if the monotone system $\mathcal{G}$ given by (4.1) has an equilibrium point at the origin, *i.e.*, $f(\mathbf{0}) + g(\mathbf{0}) = \mathbf{0}$, then it is positive.

The existence of time-delays may, in general, induce instability. However, positive monotone systems whose vector fields are *sub-homogeneous* are insensitive to *constant* delays [128]. More precisely, assume that f is cooperative and sub-homogeneous on $\mathbb{R}^n_+$ and g is order-preserving and sub-homogeneous on $\mathbb{R}^n_+$. Under these assumptions on the vector fields, if the positive system $\mathcal{G}$ without time-delays ($\tau^i_j(t) = 0$ for all

i and j) is globally asymptotically stable, then $\mathcal{G}$ is also globally asymptotically stable for any constant delays ($\tau_j^i(t) = \tau_{\max}$ for all i and j).

The main objective of this chapter is to (i) determine whether sub-homogeneous positive monotone systems are also insensitive to bounded *heterogeneous time-varying* delays; and to (ii) derive necessary and sufficient conditions for delay-independent stability of general monotone systems.

4.2 Main Results

Having established the problem formulation, we will present the main contributions of this chapter.

4.2.1 Monotone Systems

The following theorem is our first key result, which provides a sufficient condition for delay-independent stability of monotone systems, not necessarily positive, with bounded heterogeneous time-varying delays.

Theorem 4.1. *For the time-delay system $\mathcal{G}$ given by* (4.1)*, suppose that f is cooperative and g is order-preserving on $\mathcal{W}$. Suppose also that there exist two vectors w and v in $\mathcal{W}$ such that $w \le v$ and*

$$\begin{aligned} f(w) + g(w) &\ge \mathbf{0}, \\ f(v) + g(v) &\le \mathbf{0}. \end{aligned} \tag{4.4}$$

If $x^\star \in \mathcal{W}$ is the only equilibrium point of the monotone system $\mathcal{G}$ in $[w, v]$, then for all bounded heterogeneous time-varying delays, $x^\star$ is asymptotically stable with respect to initial conditions satisfying

$$w \le \varphi(t) \le v, \quad t \in [-\tau_{\max}, 0]. \tag{4.5}$$

Proof. See Appendix 4.4.3. ∎

The following example illustrates the result of Theorem 4.1.

Example 4.1. Consider the time-delay system $\mathcal{G}$ given by (4.1) with

$$f(x_1, x_2) = \begin{bmatrix} -x_1 - 1 \\ x_1 - x_2(x_2^2 - 9) + 2 \end{bmatrix}, \quad g(x_1, x_2) = \begin{bmatrix} 0 \\ x_1 \end{bmatrix}. \tag{4.6}$$

Since f is cooperative on $\mathbb{R}^n$ and g is order-preserving on $\mathbb{R}^n$, according to Proposition 4.1, the time-delay system (4.6) is monotone on $\mathbb{R}^n$. Note that as

$$f(0,0) + g(0,0) = (-1, 2) \ngeq (0,0),$$

it follows from Proposition 4.2 that the monotone system (4.6) is not positive. This system has three equilibrium points:

$$x^{\star(1)} = (-1,-3),\ \ x^{\star(2)} = (-1,0),\ \ x^{\star(3)} = (-1,3).$$

Let $w^{(1)} = (-3,-5)$ and $v^{(1)} = (1,-1)$. Clearly, $w^{(1)} \le v^{(1)}$. Since

$$f\big(w^{(1)}\big) + g\big(w^{(1)}\big) = \big(2,76\big) \ge (0,0),$$
$$f\big(v^{(1)}\big) + g\big(v^{(1)}\big) = \big(-2,-4\big) \le (0,0),$$

and $x^{\star(1)}$ is the only equilibrium point of (4.6) in $[w^{(1)}, v^{(1)}]$, it follows from Theorem 4.1 that for all bounded heterogeneous time-varying delays, $x^{\star(1)}$ is asymptotically stable with respect to initial conditions satisfying $w^{(1)} \le \varphi(t) \le v^{(1)}$, $t \in [-\tau_{\max}, 0]$. Similarly, $x^{\star(3)}$ is asymptotically stable for initial conditions satisfying $w^{(3)} \le \varphi(t) \le v^{(3)}$, $t \in [-\tau_{\max}, 0]$, where $w^{(3)} = (-3,1)$ and $v^{(3)} = (1,5)$. For example, letting $\tau_1^2(t) = 4 + \sin(t)$, $t \ge 0$, the simulation results shown in Figure 4.1 confirm that $x^{\star(1)}$ and $x^{\star(3)}$ are indeed locally asymptotically stable. ■

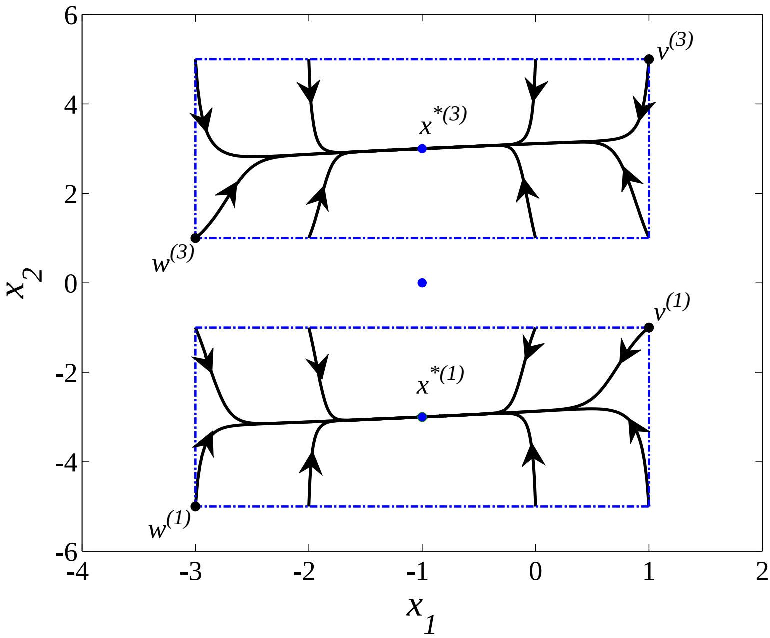


Figure 4.1: Illustration of asymptotic stability of the monotone system (4.6) in Example 4.1. The equilibrium $x^{\star(1)} = (-1,-3)$ is asymptotically stable for initial conditions $(-3,-5) \le \varphi(t) \le (1,-1)$, $t \in [-5,0]$, while $x^{\star(3)} = (-1,3)$ is asymptotically stable for initial conditions $(-3,1) \le \varphi(t) \le (1,5)$, $t \in [-5,0]$.

Remark 4.1. The delay-independent stability of monotone systems with *constant* delays was investigated in [69]. Using this result, [80, 110, 128] showed that homogeneous and sub-homogeneous positive monotone systems are insensitive to constant delays. Theorem 4.1 extends the result in [69] to bounded heterogeneous time-varying delays which include constant delays as a special case.

4.2.2 Sub-homogeneous Positive Monotone Systems

Theorem 4.1 allows us to prove that the global asymptotic stability of *positive* monotone systems whose vector fields are *sub-homogeneous* is insensitive to bounded heterogeneous time-varying delays.

Theorem 4.2. *Consider the time-delay system $\mathcal{G}$ given by (4.1). Assume that f is cooperative and g is order-preserving on $\mathbb{R}^n_+$. Furthermore, assume that f and g are sub-homogeneous of degree p with respect to the dilation map $\delta^r_\lambda(x)$. Then, the following statements are equivalent.*

(i) The sub-homogeneous positive monotone system $\mathcal{G}$ without delays ($\tau^i_j(t) = 0$ for all i and j) has a globally asymptotically stable equilibrium point at $x^\star \in \mathbb{R}^n_+$.

(ii) The sub-homogeneous positive monotone system $\mathcal{G}$ with arbitrary bounded heterogeneous time-varying delays has a globally asymptotically stable equilibrium point at $x^\star \in \mathbb{R}^n_+$.

Proof. See Appendix 4.4.4. ∎

According to Theorem 4.2, global asymptotic stability of the delay-free sub-homogeneous positive monotone system

$$\dot{x}(t) = f(x(t)) + g(x(t)), \quad t \geq 0,$$

implies that of (4.1) with bounded heterogeneous time-varying delays, and vice versa. This is a significant property of sub-homogeneous positive monotone systems since the existence of time-delays may, in general, make a stable system unstable (and, in some special cases, render an unstable system stable).

Example 4.2. Consider the time-delay system $\mathcal{G}$ given by (4.1) with

$$f(x_1, x_2) = \begin{bmatrix} -2x_1 + \frac{x_2}{x_2+2} \\ -2x_2 + \frac{x_1}{x_1+2} \end{bmatrix}, \quad g(x_1, x_2) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{4.7}$$

One can verify that f is cooperative on $\mathbb{R}^n_+$ and g is order-preserving on $\mathbb{R}^n_+$. Since $f(\mathbf{0}) + g(\mathbf{0}) = \mathbf{0}$, it follows from Proposition 4.2 that (4.7) is positive and monotone. Moreover, both f and g are sub-homogeneous of degree zero with respect to the standard dilation map. Thus, (4.7) is also sub-homogeneous. This system without time-delays has a globally asymptotically stable equilibrium at the origin [129, Example 4.1]. Therefore, Theorem 4.2 guarantees that (4.7) with any bounded time-varying delays is still globally asymptotically stable. For example, let $\tau(t) = 4+\sin(t)$, $t \geq 0$. Two state trajectories of the system starting from the initial conditions $\varphi_1(t) = (2,1)$ (solid line) and $\varphi_2(t) = (1,2)$ (dashed line), $t \in [-5,0]$, respectively, are illustrated in Figure 4.2. The origin is asymptotically stable, as expected. ■

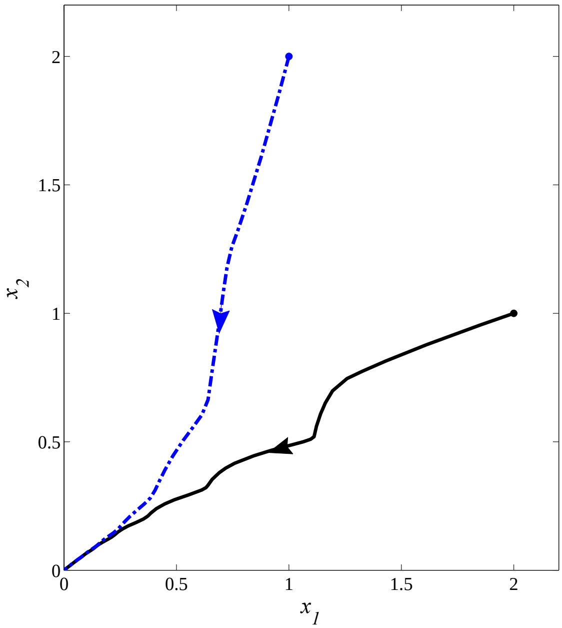


Figure 4.2: Illustration of asymptotic stability of the sub-homogeneous positive monotone system (4.7) for different initial conditions. The solid and dashed lines show trajectories of (4.7) corresponding to the initial conditions $\varphi_1(t) = (2,1)$ and $\varphi_2(t) = (1,2)$, $t \in [-5,0]$, respectively.

The next lemma, which is instrumental for the proof of Theorem 4.2, establishes a necessary condition for the global asymptotic stability of general positive monotone systems (not necessarily sub-homogeneous) with bounded heterogeneous time-varying delays.

Lemma 4.1. *For the time-delay system $\mathcal{G}$ given by* (4.1)*, suppose that f is cooperative on $\mathbb{R}^n_+$ and g is order-preserving on $\mathbb{R}^n_+$. If the positive monotone system $\mathcal{G}$ has a globally asymptotically stable equilibrium at $x^\star \in \mathbb{R}^n_+$, the following statements hold:*

(i) There does not exist a vector $w \neq x^\star$ such that $w \geq x^\star$ and

$$f(w) + g(w) \geq \mathbf{0}. \tag{4.8}$$

(ii) There exists a vector $v > \mathbf{0}$ *such that* $v > x^\star$ *and*

$$f(v) + g(v) < \mathbf{0}. \tag{4.9}$$

Proof. See Appendix 4.4.5. ∎

Lemma 4.1 provides a test for the global asymptotic stability of positive monotone systems with bounded heterogeneous time-varying delays: if we can demonstrate the existence of a vector $w \geq x^\star$ satisfying (4.8) or prove there is no positive vector $v > x^\star$ satisfying (4.9), then the equilibrium at $x^\star$ cannot be globally asymptotically stable. The following example illustrates this idea.

Example 4.3. Consider the time-delay system described by (4.1) with

$$f(x_1, x_2) = \begin{bmatrix} -x_1^2 + x_2 \\ -x_2 \end{bmatrix}, \quad g(x_1, x_2) = \begin{bmatrix} x_2 \\ x_1^2 \end{bmatrix}. \tag{4.10}$$

It is easy to verify that f is cooperative on $\mathbb{R}_+^n$, g is order-preserving on $\mathbb{R}_+^n$, and $f(0,0) + g(0,0) = (0,0)$. Thus, according to Proposition 4.2, (4.10) is a positive monotone system with an equilibrium point at the origin. Since $(1,1) \geq (0,0)$ and

$$f(1,1) + g(1,1) = (1,0) \geq (0,0),$$

it follows from Lemma 4.1 that for any bounded heterogeneous time-varying delays, the origin is not globally asymptotically stable. For example, we take $\tau_j^i(t) = 2 + \sin(t)$, $i, j = 1, 2$, $t \geq 0$, and the simulation result is shown in Figure 4.3. We can see that the trajectory of (4.10) starting from the initial condition $\varphi(t) = (1,1)$, $t \in [-3, 0]$, does not converge to the origin. ∎

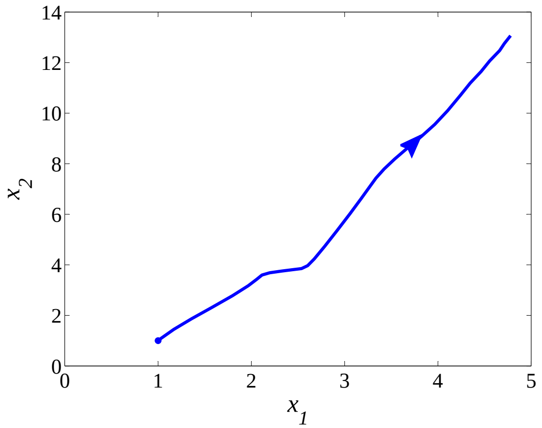


Figure 4.3: Illustration of a trajectory of the positive monotone system (4.10) corresponding to the initial condition $\varphi(t) = (1,1)$, $t \in [-3, 0]$.

Remark 4.2. Previous works in the literature established necessary conditions for global asymptotic stability of positive monotone systems without time-delays [123, 129]. Lemma 4.1, therefore, is an extension of these results to positive monotone systems of the form (4.1) with bounded heterogeneous time-varying delays.

The next example shows that the necessary conditions given in Lemma 4.1 are, in general, not sufficient for the global asymptotic stability of monotone systems.

Example 4.4. Consider the time-delay system (4.1) with

$$f(x_1, x_2) = \begin{bmatrix} -\frac{x_1}{1+x_1^3} \\ -x_2^4 \end{bmatrix}, \quad g(x_1, x_2) = \begin{bmatrix} x_2 \\ 0 \end{bmatrix}. \tag{4.11}$$

Let the time-delay be given by $\tau_2^1(t) = 5 - \cos(t)$, $t \geq 0$. It is straightforward to verify that (4.11) is a positive monotone system with an equilibrium at the origin. Since no non-zero vector $w \geq \mathbf{0}$ satisfying (4.8) exists [123, Example 3.11] and

$$f(1, \frac{1}{4}) + g(1, \frac{1}{4}) = (-\frac{1}{4}, -\frac{1}{256}) < (0, 0),$$

the necessary conditions stated in Lemma 4.1 holds. However, Figure 4.4 shows that the trajectory of (4.11) corresponding to the initial condition $\varphi(t) = (2, 1)$, $t \in [-6, 0]$, does not converge to the origin ($x_1(t)$ grows unboundedly). Therefore, the positive monotone system (4.11) is not globally asymptotically stable, which means that the necessary conditions in Lemma 4.1 are not sufficient. ∎

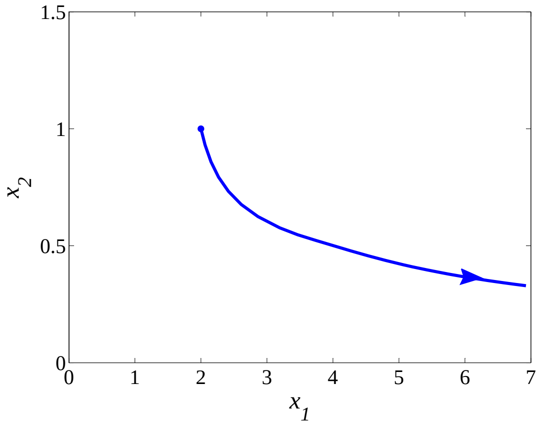


Figure 4.4: Illustration of a trajectory of the positive monotone system (4.11) corresponding to the initial condition $\varphi(t) = (2, 1)$, $t \in [-6, 0]$.

When a sub-homogeneous positive monotone system of the form (4.1) has a unique equilibrium point in $\mathbb{R}_+^n$, the necessary conditions provided by Lemma 4.1 are also sufficient. Specifically, we note the following.

Corollary 4.1. *For the time-delay system $\mathcal{G}$ given by* (4.1)*, assume that f is cooperative on $\mathbb{R}_+^n$, g is order-preserving on $\mathbb{R}_+^n$, and both f and g are sub-homogeneous of degree p with respect to the dilation map $\delta_\lambda^r(x)$. Assume also that $x^\star$ is the only equilibrium of the sub-homogeneous positive monotone system $\mathcal{G}$ in $\mathbb{R}_+^n$. Then, $\mathcal{G}$ is globally asymptotically stable for any arbitrary bounded heterogeneous time-varying delays if and only if there exists $v > \mathbf{0}$ such that $v > x^\star$ and*

$$f(v) + g(v) < \mathbf{0}.$$

The following example illustrates how the results of Corollary 4.1 can be used to analyze the delay-independent stability of a class of continuous-time power control algorithms in wireless networks.

Example 4.5 (**Continuous-time Power Control**). Consider a wireless network consisting of n mobile users. Let the transmitted power of user i at time t be given by $p_i(t)$, and define $p(t) = \big(p_1(t), \ldots, p_n(t)\big)$. A class of continuous-time power control algorithms, called *standard* power control algorithms [130], is described by

$$\dot{p}_i(t) = k_i\big(-p_i(t) + I_i(p(t))\big). \tag{4.12}$$

Here, k_i is a positive constant, $I_i : \mathbb{R}_+^n \mapsto \mathbb{R}_+$ is the *interference function* modeling the interference and noise experienced by the intended receiver of user i, and $I(p) = (I_1(p), \ldots, I_n(p))$ satisfies the following properties:

(A1) $I(p) > \mathbf{0}$.

(A2) If $p \geq p'$, then $I(p) \geq I(p')$.

(A3) For all $\alpha > 1$, $\alpha I(p) > I(\alpha p)$.

Let $K = \mathrm{diag}(k_1, \ldots, k_n)$. Then, (4.12) can be written in vector form as

$$\dot{p}(t) = K\big(-p(t) + I(p(t))\big).$$

In terms of (4.1), $x = p$, $f(x) = -Kx$ and $g(x) = KI(x)$. From (A2) and (A3), it is clear that g is order-preserving on $\mathbb{R}_+^n$ and sub-homogeneous of degree zero with respect to the standard dilation map. Since K is a diagonal matrix, $-K$ is Metzler. Thus, f is cooperative on $\mathbb{R}_+^n$. It follows from Proposition 4.1 that the standard power control algorithm (4.12) with any heterogeneous time-varying delays, given by

$$\dot{p}_i\big(t\big) = k_i\Big(-p_i(t) + I_i\big(p_1(t - \tau_1^i(t)), \ldots, p_n(t - \tau_n^i(t))\big)\Big), \tag{4.13}$$

defines a sub-homogeneous monotone system. Thanks to (A1),

$$f(\mathbf{0}) + g(\mathbf{0}) > \mathbf{0},$$

which implies that (4.13) is a positive system. Therefore, the physical constraint that the power of each user should be nonnegative ($p_i(t) \geq 0$ for all $t \geq 0$) is automatically fulfilled. The aim of this example is to study the asymptotic stability of (4.13) with bounded heterogeneous time-varying delays. Assume that the sub-homogeneous positive monotone system (4.13) has an equilibrium point in $\mathbb{R}^n_+$, *i.e.*, there exists a power vector $p^\star \in \mathbb{R}^n_+$ satisfying $I(p^\star) = p^\star$. From (A1), $p^\star > \mathbf{0}$. Note also that $p^\star$ is the unique equilibrium in $\mathbb{R}^n_+$ [126]. Pick a constant $\alpha > 1$. We have

$$\begin{aligned} f(\alpha p^\star) + g(\alpha p^\star) &= K(-\alpha p^\star + I(\alpha p^\star)) \\ &< K(-\alpha p^\star + \alpha I(p^\star)) \\ &= K(-\alpha p^\star + \alpha p^\star) \\ &= \mathbf{0}, \end{aligned}$$

where we used (A3) to get the inequality and $I(p^\star) = p^\star$ to obtain the second equality. Since $\alpha p^\star > p^\star > \mathbf{0}$ and $f(\alpha p^\star) + g(\alpha p^\star) < \mathbf{0}$, it follows from Corollary 4.1 that the standard power control algorithm (4.13) is globally asymptotically stable for all bounded heterogeneous time-varying delays (provided that (4.13) has an equilibrium point in $\mathbb{R}^n_+$). The asymptotic stability of continuous-time power control algorithms with linear interference functions was investigated for constant delays in [131] using the multivariate Nyquist criterion [132], and for bounded time-varying delays in [113] within the context of positive linear systems. In this example, since every linear interference function is also standard [126], we recover the delay independent stability of power control algorithms involving linear interference functions as a special case. ∎

4.2.3 Positive Linear Systems

We now discuss the delay-independent stability of positive linear systems of the form

$$\mathcal{G}_L : \quad \begin{cases} \dot{x}_i(t) = \sum_{j=1}^n a_{ij} x_j(t) + \sum_{j=1}^n b_{ij} x_j\big(t - \tau_j^i(t)\big), & t \geq 0, \\ x_i(t) = \varphi_i(t), & t \in [-\tau, 0], \end{cases} \tag{4.14}$$

where $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ is Metzler and $B = [b_{ij}] \in \mathbb{R}^{n\times n}$ is nonnegative. It is easy to verify that $\mathcal{G}_L$ is a sub-homogeneous positive monotone system with an equilibrium at the origin. We then have the following special case of Theorem 4.2.

Corollary 4.2. *Consider the positive linear system $\mathcal{G}_L$ given by* (4.14) *where A is Metzler and B is nonnegative. The following statements are equivalent.*

(i) $\mathcal{G}_L$ without time-delays is globally asymptotically stable.

(ii) $\mathcal{G}_L$ with any arbitrary bounded time-varying delay is globally asymptotically stable.

In the previous chapter, we have shown that the positive linear system $\mathcal{G}_L$ without time-delays is asymptotically stable if and only if $\mathcal{G}_L$ is asymptotically stable for *all* bounded time-varying delays. This result does not allow to conclude the asymptotic stability of a non-delayed positive system from the stability of the corresponding delayed system with *some* arbitrary delay (not necessarily all bounded delays). In contrast, Corollary 4.2 shows that the asymptotic stability of $\mathcal{G}_L$ under *any arbitrary* bounded time-varying delay implies the asymptotic stability of the corresponding delay-free system. Therefore, for bounded time-varying delays, Corollary 4.2 is stronger than Corollary 3.3.

4.3 Summary

We extended delay-independent stability results for sub-homogeneous positive monotone systems to allow for heterogeneous time-varying delays. Specifically, we proved that a sub-homogeneous positive monotone system with bounded heterogeneous time-varying delays is globally asymptotically stable if and only if the corresponding delay-free system is globally asymptotically stable. Homogeneous positive monotone systems and positive linear systems constitute special cases. We also derived a set of necessary and sufficient conditions for delay-independent stability of general monotone (not necessarily sub-homogeneous) systems. Illustrative examples demonstrated the validity of our theoretical results.

4.4 Appendix

In this section, we prove the main results of this chapter. We first state a key lemma which is instrumental in our argument.

4.4.1 A Technical Lemma

Lemma 4.2. *Consider the following system with constant delays, closely related to the time-delay system* (4.1) *with heterogeneous time-varying delays:*

$$\begin{cases} \dot{y}(t) = f\big(y(t)\big) + g\big(y(t - \tau_{\max})\big), & t \geq 0, \\ y(t) = \psi(t), & t \in [-\tau_{\max}, 0]. \end{cases} \tag{4.15}$$

Here, f is cooperative on $\mathcal{W}$, g is order-preserving on $\mathcal{W}$, and $\tau_{\max}$ equals the upper bound of the delays $\tau_j^i(t)$, $i, j = 1, \ldots, n$, in (4.1)*.*

1. *Assume that there exists a vector $v \in \mathcal{W}$ satisfying*

$$f(v) + g(v) \leq \mathbf{0}. \tag{4.16}$$

Assume also that the initial conditions for (4.1) *and* (4.15) *are $\varphi_v(t) = v$ and $\psi_v(t) = v$, $t \in [-\tau_{\max}, 0]$, respectively. Then, the solution $x(t, \varphi_v)$ to* (4.1) *starting from $\varphi_v(t)$ satisfies*

$$x(t, \varphi_v) \leq y(t, \psi_v), \quad \forall t \geq 0,$$

where $y(t, \psi_v)$ is the solution to (4.15) *with the initial condition $\psi_v(t)$.*

2. *Assume that there is a vector $w \in \mathcal{W}$ such that*

$$f(w) + g(w) \geq \mathbf{0}. \tag{4.17}$$

If the initial conditions for (4.1) *and* (4.15) *are $\varphi_w(t) = w$ and $\psi_w(t) = w$, $t \in [-\tau_{\max}, 0]$, respectively, then*

$$y(t, \psi_w) \leq x(t, \varphi_w), \quad \forall t \geq 0,$$

where $x(t, \varphi_w)$ and $y(t, \psi_w)$ are solutions to (4.1) *and* (4.15)*, respectively.*

Proof. **Part (1).** Let $v \in \mathcal{W}$ be a vector satisfying (4.16), and let $y(t, \psi_v)$ be the solution to (4.15) with the initial condition $\psi_v(t) = v$, $t \in [-\tau_{\max}, 0]$. Consider the following system with heterogeneous time-varying delays

$$\begin{cases} \dot{x}_i(t) = f_i\big(x(t)\big) + g_i\big(x_1(t - \tau_1^i(t)), \ldots, x_n(t - \tau_n^i(t))\big) - \frac{1}{k}, & t \geq 0, \\ x_i(t) = \varphi_i(t), & t \in [-\tau_{\max}, 0], \end{cases} \tag{4.18}$$

where $k \in \mathbb{N}$. Let $x^{(k)}(t, \varphi_v)$ be the solution to (4.18) with the initial condition $\varphi_v(t) = v$, $t \in [-\tau_{\max}, 0]$. Clearly,

$$x^{(k)}(0, \varphi_v) = v \le y(0, \psi_v) = v.$$

We claim that $x^{(k)}(t, \varphi_v) \le y(t, \psi_v)$ for all $t \ge 0$. If the result were false, we can assume that there exist an index $m \in \{1, \dots, n\}$ and a time $t_1 \ge 0$ such that

$$\begin{aligned} x^{(k)}(t, \varphi_v) &\le y(t, \psi_v), \quad \forall t \in [0, t_1], \\ x_m^{(k)}(t_1, \varphi_v) &= y_m(t_1, \psi_v), \end{aligned} \tag{4.19}$$

and

$$D^+ x_m^{(k)}(t_1, \varphi_v) \ge D^+ y_m(t_1, \psi_v). \tag{4.20}$$

Since f is cooperative, (4.19) implies that

$$f_m\big(x^{(k)}(t_1, \varphi_v)\big) \le f_m\big(y(t_1, \psi_v)\big). \tag{4.21}$$

As $t_1 - \tau_j^m(t_1) \in [-\tau_{\max}, t_1]$ for all $j \in \{1, \dots, n\}$ and $x^{(k)}(t, \varphi_v) = y(t, \psi_v) = v$ for all $t \in [-\tau_{\max}, 0]$, it follows from (4.19) that

$$x_j^{(k)}\big(t_1 - \tau_j^m(t_1), \varphi_v\big) \le y_j\big(t_1 - \tau_j^m(t_1), \psi_v\big), \tag{4.22}$$

irrespectively of whether $t_1 - \tau_j^m(t_1)$ is nonnegative or not. On the other hand, $y(t, \psi_v)$ is non-increasing for all $t \ge 0$ [69, Corollary 5.2.2]. Thus, for each j, we have

$$y_j\big(t_1 - \tau_j^m(t_1), \psi_v\big) \le y_j\big(t_1 - \tau_{\max}, \psi_v\big), \tag{4.23}$$

where we used the fact that $\tau_j^m(t_1) \le \tau_{\max}$ to get the inequality. Since g is order-preserving, it follows from (4.22) and (4.23) that

$$\begin{aligned} g_m\big(x_1^{(k)}(t_1 - \tau_1^m(t_1), \varphi_v), \dots, x_n^{(k)}(t_1 - \tau_n^m(t_1), \varphi_v)\big) \\ \le g_m\big(y_1(t_1 - \tau_{\max}, \psi_v), \dots, y_n(t_1 - \tau_{\max}, \psi_v)\big) \\ = g_m\big(y(t_1 - \tau_{\max}, \psi_v)\big). \end{aligned} \tag{4.24}$$

Using (4.21) and (4.24), the upper-right Dini-derivative of $x_m^{(k)}(t, \varphi_v)$ along the trajectories of (4.18) at $t = t_1$ satisfies

$$\begin{aligned} D^+ x_m^{(k)}(t_1, \varphi_v) &= f_m\big(x^{(k)}(t_1, \varphi_v)\big) + g_m\big(x_1(t_1 - \tau_1^m(t_1), \varphi_v), \dots, x_n(t_1 - \tau_n^m(t_1), \varphi_v)\big) - \frac{1}{k} \\ &\le f_m\big(y(t_1, \psi_v)\big) + g_m\big(y(t_1 - \tau_{\max}, \psi_v)\big) - \frac{1}{k} \\ &= D^+ y_m(t_1, \psi_v) - \frac{1}{k} \\ &< D^+ y_m(t_1, \psi_v), \end{aligned}$$

which contradicts (4.20). Therefore,

$$x^{(k)}(t, \varphi_v) \leq y(t, \psi_v), \quad \forall t \geq 0. \tag{4.25}$$

Since k was an arbitrary natural number, (4.25) holds for all $k \in \mathbb{N}$. By letting $k \to \infty$, $x^{(k)}(t, \varphi_v)$ converges to the solution $x(t, \varphi_v)$ of (4.1) uniformly on $[-\tau_{\max}, \infty)$ [99, Theorem 2.2]. This shows that $x(t, \varphi_v) \leq y(t, \psi_v)$ for all $t \geq 0$.

Part (2). Now, let $w \in \mathcal{W}$ be a vector satisfying (4.17), and let $y(t, \psi_w)$ be the solution to (4.15) with the initial condition $\psi_w(t) = w$, $t \in [-\tau_{\max}, 0]$. According to [69, Corollary 5.2.2], $y(t, \psi_w)$ is non-decreasing for all $t \geq 0$. The rest of the proof is similar to the one for Part (1) and thus omitted.

■

4.4.2 Proof of Proposition 4.2

Assume that $f(\mathbf{0}) + g(\mathbf{0}) \geq \mathbf{0}$. Let $\varphi_0(t)$ be the initial condition satisfying $\varphi_0(t) = \mathbf{0}$, $t \in [-\tau_{\max}, 0]$. Since f is cooperative on $\mathbb{R}^n_+$ and g is order-preserving on $\mathbb{R}^n_+$, it follows from Proposition 4.1 that the time-delay system (4.1) is monotone. For any nonnegative initial condition φ, we then have

$$x(t, \varphi_0) \leq x(t, \varphi), \quad \forall t \geq 0, \tag{4.26}$$

since $\varphi_0(t) \leq \varphi(t)$, $t \in [-\tau_{\max}, 0]$. Let $y(t, \psi_0)$ be the solution to the time-delay system (4.15) starting from the initial condition $\psi_0(t) = \mathbf{0}$, $t \in [-\tau_{\max}, 0]$. According to [69, Corollary 5.2.2], $y(t, \psi_0)$ is non-decreasing, *i.e*,

$$\mathbf{0} = \psi_0(0) \leq y(t, \psi_0), \quad \forall t \geq 0. \tag{4.27}$$

On the other hand, according to Lemma 4.2, $y(t, \psi_0) \leq x(t, \varphi_0)$ for all $t \geq 0$. It follows from (4.26) and (4.27) that $\mathbf{0} \leq x(t, \varphi)$ for all $t \geq 0$. Therefore, the time-delay system (4.1) is positive.

Conversely, assume that (4.1) is positive. Suppose, for contradiction, that there exists an index $m \in \{1, \ldots, n\}$ such that $f_m(\mathbf{0}) + g_m(\mathbf{0}) < 0$. Then,

$$D^+ x_m(0, \varphi_0) = f_m(\mathbf{0}) + g_m(\mathbf{0}) < 0,$$

implying that there is some $\delta > 0$ such that

$$x_m(t, \varphi_0) < x_m(0, \varphi_0) = 0, \quad \forall t \in (0, \delta).$$

Hence, $x(t) \notin \mathbb{R}^n_+$ for $t \in (0, \delta)$, which is a contradiction.

4.4.3 Proof of Theorem 4.1

Let w and v be vectors such that $w \leq v$ and that (4.4) holds. Define $\varphi_w(t) = w$ and $\varphi_v(t) = v$, $t \in [-\tau_{\max}, 0]$. Since f is cooperative and g is order-preserving, according to Proposition 4.1, the time-delay system (4.1) is monotone. Thus, for any initial condition φ satisfying (4.5), we have

$$x(t,\varphi_w) \le x(t,\varphi) \le x(t,\varphi_v), \quad \forall t \ge 0.$$

Define $\psi_w(t) = w$ and $\psi_v(t) = v$, $t \in [-\tau_{\max}, 0]$. Let $y(t,\psi_w)$ and $y(t,\psi_v)$ be solutions to the time-delay system (4.15) starting from $\psi_w(t)$ and $\psi_v(t)$, respectively. According to Lemma 4.2, $y(t,\psi_w) \le x(t,\varphi_w)$ and $x(t,\varphi_v) \le y(t,\psi_v)$ for all $t \ge 0$, implying that

$$y(t,\psi_w) \le x(t,\varphi) \le y(t,\psi_v), \quad \forall t \ge 0. \tag{4.28}$$

Since $y(t,\psi_w)$ is non-decreasing and $y(t,\psi_v)$ is non-increasing for $t \ge 0$ [69, Corollary 5.2.2], we have

$$w \le y(t,\psi_w) \le y(t,\psi_v) \le v, \quad \forall t \ge 0.$$

Thus, both $y(t,\psi_w)$ and $y(t,\psi_v)$ are bounded and monotone. It now follows from [69, Theorem 1.2.1] that $y(t,\psi_w)$ and $y(t,\psi_v)$ converge to an equilibrium of (4.15) in $[w,v]$, which must be $x^\star$, *i.e.*,

$$\lim_{t\to\infty} y(t,\psi_w) = \lim_{t\to\infty} y(t,\psi_v) = x^\star. \tag{4.29}$$

From (4.28) and (4.29), we conclude that $\lim_{t\to\infty} x(t,\varphi) = x^\star$. This completes the proof.

4.4.4 Proof of Theorem 4.2

First, we will prove that (i) implies (ii).

$(i) \Rightarrow (ii)$: Assume that the sub-homogeneous positive monotone system (4.1) without time-delays, given by

$$\dot{x}(t) = f(x(t)) + g(x(t)), \tag{4.30}$$

has a globally asymptotically stable equilibrium at $x^\star \in \mathbb{R}^n_+$. Clearly, $x^\star$ is the only equilibrium in $\mathbb{R}^n_+$. Since (4.30) is positive, according to Proposition 4.2, we have

$$f(\mathbf{0}) + g(\mathbf{0}) \ge \mathbf{0}. \tag{4.31}$$

Moreover, as $f + g$ is cooperative on $\mathbb{R}^n_+$, it follows from [129, Proposition 4.2] that there is $v > \mathbf{0}$ with $v > x^\star$ such that $f(v) + g(v) < \mathbf{0}$. This together with sub-homogeneity of f and g implies that for any real constant γ with $\gamma \ge 1$, we have

$$f(\delta^r_\gamma(v)) + g(\delta^r_\gamma(v)) \le \gamma^p \delta^r_\gamma\big(f(v) + g(v)\big) < \mathbf{0}. \tag{4.32}$$

For the time-delay system (4.1), it follows from Theorem 4.1 and inequalities (4.31) and (4.32) that $x^\star$ is asymptotically stable when the initial condition φ satisfies

$$\mathbf{0} \le \varphi(t) \le \delta^r_\gamma(v), \quad t \in [-\tau_{\max}, 0]. \tag{4.33}$$

Let $\varphi \in \mathcal{C}\big([-\tau_{\max}, 0], \mathbb{R}^n_+\big)$ be an arbitrary nonnegative initial condition. As $\varphi(t)$ is continuous and, hence, bounded on $[-\tau_{\max}, 0]$, we can find $\gamma \geq 1$ sufficiently large such that (4.33) holds. Therefore, the time-delay system (4.1) with arbitrary bounded heterogeneous time-varying delays is globally asymptotically stable for all nonnegative initial conditions.

We next show that (ii) implies (i).

$(ii) \Rightarrow (i)$: Assume that the time-delay system (4.1) with arbitrary bounded heterogeneous time-varying delays $\tau^i_j(t)$, $i, j = 1, \ldots, n$, has a globally asymptotically stable equilibrium at $x^\star \in \mathbb{R}^n_+$. Since (4.1) is positive, it follows from Proposition 4.2 that (4.31) holds. Moreover, according to Lemma 4.1, there exists a vector $v > \mathbf{0}$ such that $v > x^\star$ and satisfies (4.9). Let $x(t, x_0)$ be the solution of the non-delayed system (4.30) corresponding to the initial condition $x(0) = x_0$. The proof will broken up into three steps:

1. First, we show that the solutions $x(t, \mathbf{0})$ and $x(t, v)$ of (4.30) starting from the initial conditions $x(0) = \mathbf{0}$ and $x(0) = v$, respectively, converge to $x^\star$ as $t \to \infty$.

2. Second, we prove that for any nonnegative initial condition $x_0 \in \mathbb{R}^n_+$, there exists a vector $\overline{v} > \mathbf{0}$ such that $\overline{v} \geq x_0$ and that (4.9) holds for $\overline{v}$.

3. Finally, we show that for any $x_0 \in \mathbb{R}^n_+$, the solution $x(t, x_0)$ of (4.30) converges to $x^\star$ as $t \to \infty$.

Step 1. Since $f + g$ is cooperative, it follows from [69, Proposition 3.2.1] that $x(t, \mathbf{0})$ is non-decreasing and $x(t, v)$ is non-increasing for all $t \geq 0$, which implies that

$$\mathbf{0} \leq x(t, \mathbf{0}) \leq x(t, v) \leq v, \quad t \geq 0.$$

Thus, $x(t, \mathbf{0})$ and $x(t, v)$ are bounded and monotone. It now follows from [69, Theorem 1.2.1] that $x(t, \mathbf{0})$ and $x(t, v)$ converge to an equilibrium of (4.30) in $[0, v]$, which means that $x(t, \mathbf{0}), x(t, v) \to \bar{x}^\star$, where

$$f(\bar{x}^\star) + g(\bar{x}^\star) = \mathbf{0}.$$

We claim that $\bar{x}^\star = x^\star$. By contradiction, suppose this is not true. Then, it is easy to verify that

$$x(t) = \bar{x}^\star \neq x^\star, \quad t \in [-\tau_{\max}, \infty),$$

satisfies (4.1). This shows that for the nonnegative initial condition $\varphi_{\bar{x}^\star}(t) = \bar{x}^\star$, $t \in [-\tau_{\max}, 0]$, the solution $x(t, \varphi_{\bar{x}^\star})$ of (4.1) does not converge to $x^\star$, contradicting the fact that $x^\star$ is the globally asymptotically stable equilibrium of (4.1). Therefore, $x(t, \mathbf{0})$ and $x(t, v)$ converge to $x^\star$ as $t \to \infty$.

Step 2. Let $x_0 \in \mathbb{R}^n_+$ be an arbitrary initial condition and let $v > \mathbf{0}$ be a vector satisfying (4.9). Then, we can choose $\gamma \geq 1$ such that $x_0 \leq \delta^r_\gamma(v)$. Define $\overline{v} = \delta^r_\gamma(v)$. As f and g are sub-homogeneous, we have

$$f(\overline{v}) + g(\overline{v}) \leq \gamma^p \delta^r_\gamma\big(f(v) + g(v)\big) < \mathbf{0},$$

where the right-most inequality follows from (4.9).

Step 3. According to the previous step, for any initial condition $x_0 \in \mathbb{R}^n_+$, we can find $\overline{v} > \mathbf{0}$ such that $\overline{v} \geq x_0$ and that (4.9) holds for $\overline{v}$. As $f + g$ is cooperative, the non-delayed system (4.30) is monotone [69, p. 34], implying that

$$x(t, \mathbf{0}) \leq x(t, x_0) \leq x(t, \overline{v}), \quad \forall t \geq 0.$$

From the first step, we have $x(t, \mathbf{0}), x(t, \overline{v}) \to x^\star$ as $t \to \infty$. Therefore, $x(t, x_0)$ also converges to $x^\star$. This shows that (4.30) is globally asymptotically stable for all nonnegative initial conditions.

4.4.5 Proof of Lemma 4.1

We first show that if the time-delay system (4.1) has a globally asymptotically stable equilibrium at $x^\star \in \mathbb{R}^n_+$, then (i) holds.

(i) Assume that there is $w \neq x^\star$ such that $w \geq x^\star$ and that (4.8) holds. Define $\varphi_w(t) = w$ and $\psi_w(t) = w$, $t \in [-\tau_{\max}, 0]$. According to Lemma 4.2, we have

$$y(t, \psi_w) \leq x(t, \varphi_w), \quad \forall t \geq 0, \tag{4.34}$$

where $x(t, \varphi_w)$ and $y(t, \psi_w)$ are solutions of (4.1) and (4.15), respectively. Moreover, $y(t, \psi_w)$ is non-decreasing for all $t \geq 0$ [69, Corollary 5.2.2], implying that

$$w \leq y(t, \psi_w), \quad \forall t \geq 0. \tag{4.35}$$

Therefore, from (4.34) and (4.35), we have $w \leq x(t, \varphi_w)$ for all $t \geq 0$. This means that $x(t, \varphi_w) \nrightarrow x^\star$, contradicting the fact that $x^\star$ is globally asymptotically stable.

(ii) According to part (i), we have

$$(f + g)(w) \ngeq \mathbf{0}, \quad \forall w \geq x^\star,\ w \neq x^\star.$$

Since $f + g$ is cooperative on $\mathbb{R}^n_+$, it follows from [129, Proposition 4.2] that there is $v > x^\star$ satisfying (4.9).

Chapter 5

Asynchronous Contractive Iterations

Asynchrnous iterations appear naturally in parallel and distributed systems and are heavily exploited in applications ranging from large-scale linear algebra and optimization to distributed coordination of small embedded devices (see, e.g., [133–137] and references therein). Allowing nodes to operate in an asynchronous manner simplifies the implementation of distributed algorithms and eliminates the overhead associated with synchronization. However, care has to be taken since asynchrony runs the risk of rendering an otherwise stable iteration unstable.

The dynamics of asynchronous iterations are much richer than their synchronous counterparts, and quantifying the impact of asynchrony on the convergence properties of iterative algorithms remains challenging. Some of the first results on the convergence of asynchronous iterations were derived by Chazan and Miranker [138], who studied chaotic relaxations for solving linear systems of equations. Several authors have proposed extensions of this pioneering work to nonlinear iterations involving maximum norm pseudo-contractions (e.g., [139–141]) and to monotone iterations (e.g., [142–144]). Powerful convergence results for broad classes of asynchronous iterations, including maximum norm pseudo-contractions and monotone mappings, under different assumptions on communication delays and update rates were presented by Bertsekas [145] and Bertsekas and Tsitsiklis [16]. However, most of the results in the literature only guarantee *asymptotic* convergence of asynchronous iterations. This chapter complements the existing work by developing convergence theorems that characterize *convergence rates* of asynchronous iterations and *quantify* how these rates depend on the update intervals and information delays.

Contributions of the Chapter. We focus on iterations involving maximum norm pseudo-contractions under the general asynchronous model introduced in [16, 145], which allows for heterogeneous and time-varying communication delays and update rates. These iterations arise in a variety of algorithms, such as certain classes of linear fixed-point iterations and gradient descent methods [16, 146], optimum multiuser detection algorithms [147], distributed algorithms for averaging [148], and power control algorithms in wireless networks [14]. Our main theorem provides a powerful approach for characterizing the rate of convergence of totally asynchronous iterations, where both the update intervals and communication delays may grow unbounded. When specialized to partially asynchronous iterations (where the update intervals and communication delays have a fixed upper bound), or to particular classes of unbounded delays and update intervals, our approach allows to explicitly quantify how the degree of asynchronism affects the convergence rate. Several examples are worked out to demonstrate that our main theorem recovers and improves on existing results, and that it allows to characterize the solution times for several classes of asynchronous iterations that have not been addressed before.

Outline of the Chapter. Section 5.1 reviews partially and totally asynchronous models of computation and recalls some basic results about fixed point iterations involving pseudo-contractions in the block-maximum norm. Section 5.2 presents our main results on the convergence rates of asynchronous iterations. Section 5.3 demonstrates how our results can be used to analyze the impact of asynchronism on the convergence rate of power control algorithms in wireless networks. Numerical examples are presented in Section 5.4 to illustrate the accuracy of our guaranteed bounds on the convergence rate of asynchronous power control algorithms. Finally, concluding remarks are given in Section 5.5.

5.1 Problem Formulation

We consider iterative algorithms on the form

$$x_i\big(t+1\big) = f_i\big(x_1(t), \dots, x_m(t)\big), \quad t \in \mathbb{N}_0, \tag{5.1}$$

where $i = 1, \dots, m$, $x_i \in \mathbb{R}^{n_i}$, and $f_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$ are continuous functions with $n = n_1 + \dots + n_m$. A vector $x^\star = (x_1^\star, \dots, x_m^\star) \in \mathbb{R}^n$ is called a *fixed point* of the function $f(x) = (f_1(x), \dots, f_m(x))$ if

$$x_i^\star = f_i(x_1^\star, \dots, x_m^\star), \quad i = 1, \dots, m.$$

The iteration (5.1) can be viewed as a network of m nodes, each responsible for updating one of the m subvectors of x so as to find a global fixed point of f. The reason is that when the sequence $\{x_i(t)\}$ generated by (5.1) converges to $x_i^\star$ for each i, then $x^\star$ is a fixed point of f.

For spaces $\mathbb{R}^{n_i}$, let us fix some norms $\|\cdot\|_i$, $i = 1, \dots, m$. Then, for each vector $x = (x_1, \dots, x_m) \in \mathbb{R}^n$, the *block-maximum norm* is defined by

$$\|x\|_b^w = \max_{1 \le i \le m} \frac{\|x_i\|_i}{w_i},$$

where w_i are positive constants. Note that when $n_i = 1$ for each i, the block-maximum norm reduces to the maximum norm defined by

$$\|x\|_\infty^w = \max_{1 \le i \le m} \frac{|x_i|}{w_i}.$$

Definition 5.1. A mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ is called a *pseudo-contraction* with respect to the block-maximum norm if there exists $c \in [0, 1)$ such that

$$\|f(x) - x^\star\|_b^w \le c\, \|x - x^\star\|_b^w, \quad \forall x \in \mathbb{R}^n,$$

where $x^\star$ is a fixed point of f. The scalar c is called the *contraction modulus* of f.

Pseudo-contractions have at most one fixed point, to which the iterates produced by (5.1) converge at a linear rate [16]. Note that in contrast to contraction mappings, the existence of a fixed point is part of the definition of pseudo-contractions.

The algorithm described by (5.1) is *synchronous* in the sense that all nodes update their states at the same time and have access to the states of all other nodes. Synchronous execution is possible if there are no communication faults or delays in the network and all nodes operate in synch with a global clock. In practice, these requirements are hard to satisfy: local clocks in different nodes tend to drift and the communication latency between nodes can be significant and unpredictable. The drawback with insisting on synchronous operation in an inherently asynchronous environment is that nodes will spend a significant time idle, especially if some nodes are faster because of, for example, higher processor power or smaller workload per iteration.

In an *asynchronous* implementation of the iteration (5.1), each node updates its state at its own pace, using possibly outdated information from the other nodes. Such iterations can be written as

$$x_i(t+1) = \begin{cases} f_i\big(x_1(s_1^i(t)), \dots, x_m(s_m^i(t))\big), & t \in T^i, \\ x_i(t), & t \notin T^i, \end{cases} \tag{5.2}$$

where T^i is the set of times when node i executes an update, and $s_j^i(t)$ is the time at which the most recent version of x_j available to node i at time t was computed [16]. Note that $0 \le s_j^i(t) \le t$ for all $t \in \mathbb{N}_0$. We can view

$$\tau_j^i(t) := t - s_j^i(t),$$

as the communication delay from node j to node i at time t. It is clear that the synchronous iteration (5.1) is a special case of (5.2) where $\tau_j^i(t) = 0$, and $T^i = \mathbb{N}_0$ for all $t \in \mathbb{N}_0$ and all $i, j = 1, \dots, m$.

Based on the assumptions on communication delays and update rates, asynchronous algorithms are classified into *totally asynchronous* and *partially asynchronous*:

Assumption 5.1 (**Total Asynchronism [16, §6]**). *For the asynchronous iteration* (5.2)*, there holds:*

1. *The sets T^i are infinite subsets of $\mathbb{N}_0$ for all $i = 1, \dots, m$.*

2. $\lim_{t\to\infty} s_j^i(t) = \infty$ *for all $i, j = 1, \dots, m$.*

Loosely speaking, Assumption 5.1.1 guarantees that no node ceases to execute its update while Assumption 5.1.2 guarantees that old information is eventually purged from the computation. Under total asynchronism, the delay $\tau_j^i(t)$ can become unbounded as t increases. This is the main difference with partially asynchronous iterations, where delays are bounded; in particular, the following assumption holds.

Assumption 5.2 (**Partial Asynchronism [16, §7]**). *For the asynchronous iteration* (5.2)*, there exists a positive integer B such that:*

1. *For every i and for every $t \in \mathbb{N}_0$, at least one of the elements of the set $\{t, t+1, \dots, t+B-1\}$ belongs to T^i.*
2. *There holds $0 \leq \tau_j^i(t) \leq B-1$, for all i and j, and all $t \in \mathbb{N}_0$ belonging to T^i.*
3. *There holds $\tau_i^i(t) = 0$ for all i and $t \in T^i$.*

Assumptions 5.2.1 and 5.2.2 ensure that both the communication delays and the time interval between updates executed by each node are bounded. When $B = 1$, this model reduces to the synchronous iteration (5.1). Assumption 5.2.3 states that node i always uses the latest version of its own component x_i.

While convergent synchronous iterations may diverge in the face of asynchronism, the asynchronous iteration (5.2) involving pseudo-contractions in the block-maximum norm also converges to the fixed point under total asynchronism [16]. In other words, if f is a pseudo-contraction with respect to the block-maximum norm, then the synchronous iteration (5.1) can tolerate arbitrary large communication and computation delays. However, [16] did not quantify how bounds on the time-delays and update rates of nodes affect the convergence rate of (5.2). One could expect that the convergence rate would become slower with increasing communication delays or with more infrequent updates. Our main objective in this chapter is therefore to give explicit estimates of the convergence rate of asynchronous iterations involving block-maximum norm pseudo-contractions under different assumptions on communication delays and update rates.

5.2 Convergence Rate of Asynchronous Iterations

We will now develop a theorem that provides guaranteed convergence rates of the asynchronous iteration (5.2) under various classes of total asynchronism. Our proof uses a continuous non-increasing function $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying

$$\lim_{t \to \infty} \lambda(t) = 0,$$

and shows that for all $i = 1, \dots, m$, and for all $t \in \mathbb{N}_0$,

$$\frac{1}{w_i} \|x_i(t) - x_i^\star\|_i \leq M\lambda(t_k^i), \quad t \in (t_k^i, t_{k+1}^i],$$

where M is a positive constant, and t_k^i and t_{k+1}^i are two consecutive elements of the set T^i. The function $\lambda(t)$ quantifies how fast the sequence of vectors generated by (5.2) converges to the fixed point $x^\star$. For example, if $\lambda(t) = \rho^t$ with $\rho \in (0,1)$, then $\{x_i(t_k^i)\}$ converges at a linear rate to $x_i^\star$; and if $\lambda(t) = t^{-\xi}$ with $\xi > 0$, then $\|x_i(t_k^i) - x_i^\star\|_i$ is upper bounded by a polynomial function of time. Similar to the asynchronous iterates themselves, the guaranteed upper bound is left unchanged when $t \notin T^i$ and decreases after update times; see Figure 5.1.

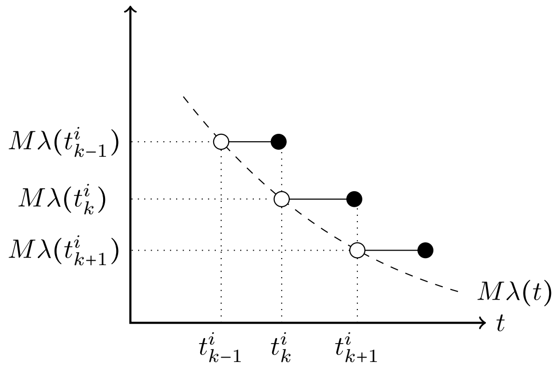


Figure 5.1: Illustration of the guaranteed upper bound on the iterates generated by the asynchronous iteration (5.2). Here, t_{k-1}^i, t_k^i and t_{k+1}^i are three consecutive elements of the set T^i.

Theorem 5.1. *Consider the asynchronous iteration* (5.2)*. Suppose that the following conditions hold:*

i) *f is a pseudo-contraction with respect to the block-maximum norm with contraction modulus c.*

ii) *There exist $\Delta \in \mathbb{N}_0$ and functions $\beta^i : \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $t \geq \Delta$,*

$$t - t_k^i \leq \beta^i(t) \leq t, \quad t \in (t_k^i, t_{k+1}^i], \tag{5.3}$$

where t_k^i and t_{k+1}^i are two consecutive elements of T^i.

iii) *There is a non-increasing function $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$\lim_{t\to\infty} \lambda(t) = 0,$$

and that for all i and j,

$$\lim_{t\to\infty} \frac{\lambda\big(s_j^i(t) - \beta^j(s_j^i(t))\big)}{\lambda(t)} < \frac{1}{c}. \tag{5.4}$$

Then, for all i and all $t \in \mathbb{N}$, the sequence of vectors generated by (5.2) *under total asynchronism satisfies*

$$\frac{1}{w_i}\|x_i(t) - x_i^\star\|_i \le M\lambda(t_k^i), \quad t \in (t_k^i, t_{k+1}^i],$$

where M is a positive constant.

Remark 5.1. The value $\beta^i(t_{k+1})$ is an upper bound on the time interval between node i's k^{th} and $k+1^{\text{st}}$ updates. Letting $\beta^i(t) = \beta$, $\beta \in \mathbb{N}$, means that node i performs at least one update during any time interval of length β. In general, $\beta^i(t)$ may be unbounded (we will consider such a case in Example 1).

We are now ready to prove the theorem.

Proof. (*of Theorem 5.1*)
For each $i = 1, \dots, m$, let t_0^i be the first element of the set T^i, *i.e.*, $t_k^i \ge t_0^i$ for all $t_k^i \in T^i$. According to Assumption 5.1.2, there exists a time $\widehat{t} \in \mathbb{N}_0$ large enough such that for all i and j,

$$s_j^i(t) \ge \max\Big\{\Delta, \max_{1\le i\le m}\{t_0^i\} + 1\Big\}, \quad \forall t \ge \widehat{t}. \tag{5.5}$$

From (5.4), we can find a sufficiently large time $\widetilde{t} \in \mathbb{N}_0$ so that

$$c\lambda\big(s_j^i(t) - \beta^j(s_j^i(t))\big) \le \lambda\big(t\big), \quad \forall t \ge \widetilde{t}. \tag{5.6}$$

Let $\bar{t} = \max\{\,\widehat{t}, \widetilde{t}\,\}$, and define

$$M = \frac{\|x(0) - x^\star\|_b^w}{\lambda(\bar{t})}.$$

It follows from Proposition 2.1 in [16, §6.2] that the sequence $\{x(t)\}$ generated by the asynchronous iteration (5.2) satisfies

$$\frac{1}{w_i}\|x_i(t) - x_i^\star\|_i \le \|x(0) - x^\star\|_b^w, \quad \forall t \in \mathbb{N}_0,$$

for each $i = 1, \dots, m$. Thus,

$$\begin{aligned}\max_{0\le t\le \bar{t}}\left\{\frac{1}{w_i}\frac{\|x_i(t) - x_i^\star\|_i}{\lambda(t)}\right\} &\le \max_{0\le t\le \bar{t}}\left\{\frac{\|x(0) - x^\star\|_b^w}{\lambda(t)}\right\}\\ &\le \frac{\|x(0) - x^\star\|_b^w}{\lambda(\bar{t})}\\ &= M,\end{aligned}$$

where for the second inequality, we used the fact that $\lambda(t)$ is non-increasing on $\mathbb{R}_+$. This implies that

$$\frac{1}{w_i}\|x_i(t) - x_i^\star\|_i \leq M\lambda(t),$$

for $t \in \{0, \dots, \bar{t}\}$. Since $\lambda(t)$ is non-increasing, we have $\lambda(t) \leq \lambda(t_k^i)$ for $t \geq t_k^i$. Thus,

$$\frac{1}{w_i}\|x_i(t) - x_i^\star\|_i \leq M\lambda(t_k^i), \quad t \in (t_k^i, t_{k+1}^i], \tag{5.7}$$

for all $t \in \{0, \dots, \bar{t}\}$. We will show by induction that (5.7) also holds for all $t \geq \bar{t}$. Assume for induction that (5.7) holds for all t up to some $t' \in \mathbb{N}_0$, where $t' \geq \bar{t}$. Let $t_{k'}^i$ and $t_{k'+1}^i$ be two consecutive elements of T^i such that $t' \in (t_{k'}^i, t_{k'+1}^i]$. Using the induction hypothesis, we have

$$\frac{1}{w_i}\|x_i(t') - x_i^\star\|_i \leq M\lambda(t_{k'}^i). \tag{5.8}$$

We now prove that $x_i(t'+1)$ satisfies (5.7).

Case 1) If $t' \notin T^i$, then $t'+1 \in (t_{k'}^i, t_{k'+1}^i]$. Moreover, from (5.2), $x_i(t'+1) = x_i(t')$. It follows from (5.8) that

$$\frac{1}{w_i}\|x_i(t'+1) - x_i^\star\|_i = \frac{1}{w_i}\|x_i(t') - x_i^\star\|_i \leq M\lambda(t_{k'}^i).$$

Therefore, in this case, (5.7) is true for $t'+1$.

Case 2) If $t' \in T^i$, or, equivalently, $t' = t_{k'+1}^i$, then

$$\begin{aligned}\frac{1}{w_i}\|x_i(t'+1) - x_i^\star\|_i &= \frac{1}{w_i}\left\|f_i\big(x_1(s_1^i(t')), \dots, x_m(s_m^i(t'))\big) - x_i^\star\right\|_i \\ &\leq c\left\|\big(x_1(s_1^i(t')), \dots, x_m(s_m^i(t'))\big) - x^\star\right\|_b^w \\ &= c \max_{1\leq j\leq m}\left\{\frac{1}{w_j}\|x_j(s_j^i(t')) - x_j^\star\|_j\right\},\end{aligned} \tag{5.9}$$

where the inequality holds since f is a pseudo-contraction with respect to the block-maximum norm. As $t' \geq \bar{t} \geq \widehat{t}$, (5.5) implies that $s_j^i(t') > t_0^j$ for each j. Let $t_{k_s}^j$ and $t_{k_s+1}^j$ be two consecutive elements of T^j such that

$$s_j^i(t') \in (t_{k_s}^j, t_{k_s+1}^j].$$

Since $s_j^i(t') \leq t'$, the induction hypothesis yields

$$\frac{1}{w_j}\|x_j(s_j^i(t')) - x_j^\star\|_i \leq M\lambda(t_{k_s}^j), \tag{5.10}$$

for all j. Moreover, (5.5) also implies that $s_j^i(t') \geq \Delta$. It follows from (5.3) that

$$t_{k_s}^j \geq s_j^i\big(t'\big) - \beta^j\big(s_j^i(t')\big) \geq 0.$$

As $\lambda(t)$ is non-increasing on $\mathbb{R}_+$, this in turn implies

$$\lambda\big(t_{k_s}^j\big) \leq \lambda\big(s_j^i(t') - \beta^j(s_j^i(t'))\big). \tag{5.11}$$

Substituting (5.10) into (5.9), then using (5.11), we obtain

$$\begin{aligned}\frac{1}{w_i}\|x_i(t'+1) - x_i^\star\|_i &\leq cM \max_{1\leq j\leq m} \lambda(t_{k_s}^j) \\ &\leq cM \max_{1\leq j\leq m} \lambda\big(s_j^i(t') - \beta^j(s_j^i(t'))\big) \\ &\leq M\lambda(t') \\ &= M\lambda(t_{k'+1}^i),\end{aligned} \tag{5.12}$$

where the last inequality follows from (5.6). Note that

$$t' + 1 = t_{k'+1}^i + 1 > t_{k'+1}^i,$$

implying that $t'+1 \in (t_{k'+1}^i, t_{k'+2}^i]$. From (5.12), we conclude that (5.7) holds for $t'+1$. The induction proof is complete. ∎

Theorem 5.1 shows that any function $\lambda(t)$ satisfying condition (iii) can be used to estimate the convergence rate of totally asynchronous iterations. From (5.4), it is clear that the admissible choices for $\lambda(t)$ depend on the asymptotic behaviour of $\beta^i(t)$ and $s_j^i(t)$. This means that the rate at which the nodes execute their updates as well as the way communication delays tend large affect the convergence rate. To clarify this statement, we will analyze a few special cases in detail. First, we consider the partially asynchronous model. The following result gives a bound on the convergence rate of asynchronous iterations involving block-maximum norm pseudo-contractions under this model of asynchronicity.

Theorem 5.2. *Consider the asynchronous iteration* (5.2) *under partial asynchronism. Assume that f is a block-maximum norm pseudo-contraction with contraction modulus c. Then, for all i and all $t \in \mathbb{N}$, we have*

$$\frac{1}{w_i}\|x_i(t) - x_i^\star\|_i \leq M\rho^{t_k^i}, \quad t \in (t_k^i, t_{k+1}^i], \tag{5.13}$$

where M is a positive constant, t_k^i and t_{k+1}^i are two consecutive elements of T^i, and

$$\rho = c^{\frac{1}{2B-1}}. \tag{5.14}$$

Proof. According to Assumption 5.2.1, we have

$$t - t_k^i \le B \le t, \quad t \in (t_k^i, t_{k+1}^i],$$

for $t \ge B$. Thus, we can choose $\beta^i(t) = B$, $i = 1, \dots, m$. Pick a constant $\hat{\rho}$ such that

$$\hat{\rho} \in (\rho, 1), \tag{5.15}$$

where ρ is defined by (5.14). Let $\lambda(t) = \hat{\rho}^t$, $t \ge 0$. Clearly, $\lambda(t)$ is non-increasing on $\mathbb{R}_+$. Moreover, for all i and j, we obtain

$$\begin{aligned} c \lim_{t\to\infty} \frac{\lambda\big(s_j^i(t) - \beta^j(s_j^i(t))\big)}{\lambda(t)} = c \lim_{t\to\infty} \frac{\hat{\rho}^{s_j^i(t)-B}}{\hat{\rho}^t} &= c \lim_{t\to\infty} \frac{\hat{\rho}^{t-\tau_j^i(t)-B}}{\hat{\rho}^t} \\ &\le c \lim_{t\to\infty} \frac{\hat{\rho}^{t+1-2B}}{\hat{\rho}^t} \\ &= c\hat{\rho}^{1-2B} \\ &< c\rho^{1-2B} \\ &= 1. \end{aligned}$$

The first inequality uses the fact that under Assumption 5.2.2, $\tau_j^i(t) \le B - 1$ for $t \in \mathbb{N}_0$. The last equality uses (5.14). Therefore, condition (iii) of Theorem 5.1 holds for all $\hat{\rho}$ satisfying (5.15). Hence, the sequence $\{x(t)\}$ generated by the asynchronous iteration (5.2) satisfies (5.13).

■

According to Theorem 5.2, asynchronous iterations involving block-maximum norm pseudo-contractions still converge at a linear rate under partial asynchronism. Note that $c^{1/(2B-1)}$ is monotonically increasing with B and approaches one as B tends to infinity. Hence, the guaranteed convergence rate of partially asynchronous iterations deteriorates with increasing delays.

Contrary to the typical upper bounds on the convergence rate, the guaranteed bounds provided by Theorem 5.1 do not decrease at every time step, but only at the update times $t_k^i \in T^i$. Therefore, our estimation of convergence rate, in general, depends on how fast the sequence $\{t_k^i\}$ grows large. For example, Theorem 5.2 shows that the sequence $\{\|x_i(t) - x_i^\star\|_i\}$ generated by the partially asynchronous iteration (5.2) is upper bounded by $M\rho^{t_k^i}$. However, under partial asynchronism, we have

$$0 \le t - B \le t_k^i, \quad t \in (t_k^i, t_{k+1}^i],$$

for all $t \ge B$. Thus,

$$M\rho^{t_k^i} \le M\rho^{t-B} = M'\rho^t, \quad t \in (t_k^i, t_{k+1}^i],$$

where $M' = M\rho^{-B}$. This shows that

$$\frac{1}{w_i}\|x_i(t) - x_i^\star\|_i \le M'\rho^t,$$

for each i and $t \ge B$. Therefore, partially asynchronous iterations involving block-maximum norm pseudo-contractions attains a rate of $O(\rho^t)$.

Under partial asynchronism, both update rates and communication delays are bounded. However, Theorem 5.1 can also be used to find guaranteed convergence rates of asynchronous iterations with unbounded communication delays and update intervals. To make our point, we establish convergence rates for a particular class of totally asynchronous iterations described by the following assumption:

Assumption 5.3. *For the asynchronous iteration* (5.2)*, there exist* $B \in \mathbb{N}$*,* $\alpha \in [0, 1)$*, and* $t_\alpha \in \mathbb{N}_0$ *such that*

1. *For every* i *and for every* $t \in \mathbb{N}_0$*, at least one of the elements of the set* $\{t, t+1, \dots, t+B-1\}$ *belongs to* T^i*.*
2. $0 \le \tau_j^i(t) \le \alpha t$*, for all* $i, j = 1, \dots, m$*, and all* $t \ge t_\alpha$*.*

Time-delays satisfying Assumption 5.3.2 may be unbounded (take, for example, $\tau_j^i(t) = \lfloor 0.2t \rfloor$, $t \in \mathbb{N}_0$). The associated convergence result now reads as follows.

Theorem 5.3. *Consider the asynchronous iteration* (5.2) *under Assumption 5.3. Assume that* f *is a pseudo-contraction with contraction modulus* c *with respect to the block-maximum norm. Then, the sequence* $\{x(t)\}$ *generated by* (5.2) *satisfies*

$$\frac{1}{w_i}\|x_i(t) - x_i^\star\|_i \le M\left(\frac{t_k^i}{B} + 1\right)^{-\xi}, \quad t \in (t_k^i, t_{k+1}^i], \tag{5.16}$$

where M *is a positive constant,* t_k^i *and* t_{k+1}^i *are two consecutive elements of* T^i*, and*

$$\xi = \frac{\ln c}{\ln(1-\alpha)}. \tag{5.17}$$

Proof. Similar to the proof of Theorem 5.2, we choose $\beta^i(t) = B$, $i = 1, \dots, m$. Let

$$\lambda(t) = \left(\frac{t}{B} + 1\right)^{-\widehat{\xi}}, \quad t \ge 0,$$

where $\widehat{\xi}$ is a positive constant satisfying

$$\widehat{\xi} \in (0, \xi). \tag{5.18}$$

We then have

$$
\begin{aligned}
c \lim_{t\to\infty} \frac{\lambda\big(s_j^i(t) - \beta^j(s_j^i(t))\big)}{\lambda(t)} = c \lim_{t\to\infty} \left(\frac{t/B+1}{(s_j^i(t)-B)/B+1}\right)^{\widehat{\xi}} &= c \lim_{t\to\infty} \left(\frac{t+B}{t-\tau_j^i(t)}\right)^{\widehat{\xi}} \\
&\le c \lim_{t\to\infty} \left(\frac{t+B}{(1-\alpha)t}\right)^{\widehat{\xi}} \\
&= \frac{c}{(1-\alpha)^{\widehat{\xi}}} \\
&< \frac{c}{(1-\alpha)^{\xi}} \\
&= 1,
\end{aligned}
$$

where for the first inequality, we used the fact that

$$
0 \le (1-\alpha)t \le t - \tau_j^i(t), \quad t \ge t_\alpha.
$$

The second inequality follows from (5.18). Therefore, according to Theorem 5.1, the sequence $\{x(t)\}$ generated by the asynchronous iteration (5.2) satisfies (5.16).

■

Theorem 5.3 shows that the convergence rate of the asynchronous algorithm (5.2) under unbounded delays satisfying Assumption 5.3 is upper bounded by a polynomial function of time. From (5.17), we can see that the rate at which the unbounded delays grow large, α, affects ξ. Specifically, ξ is monotonically decreasing with α and approaches zero as α tends to one. In addition, the upper bound on the convergence rate is inversely proportional to B. It follows that the guaranteed convergence rates get increasingly slower as either delays are allowed to grow quicker when $t \to \infty$ or nodes execute less frequently.

Remark 5.2. Under Assumption 5.3.1, we have

$$
0 \le t - B \le t_k^i, \quad t \in (t_k^i, t_{k+1}^i],
$$

for all $t \ge B$, implying that $t/B \le t_k^i/B + 1$. Therefore, according to Theorem 5.3, the sequence $\{x(t)\}$ generated by (5.2) satisfies

$$
\frac{1}{w_i} \|x_i(t) - x_i^\star\|_i \le M \left(\frac{t}{B}\right)^{-\xi}, \quad t \ge B.
$$

This shows that the convergence rate of the asynchronous iteration (5.2) with unbounded delays satisfying Assumption 5.3 is $O(t^{-\xi})$.

The guaranteed bounds provided by Theorems 5.2 and 5.3 are derived under the assumption that the update intervals of all nodes are bounded by a constant B, *i.e.*,

$$
t_{k+1}^i - t_k^i \le B, \tag{5.19}
$$

for any two consecutive elements of T^i. However, Theorem 5.1 allows time-varying upper bounds on both update rates and communication delays. Rather than developing theorems for specific combinations of update rates and time-delays, we illustrate the principle on a simple example.

Example 5.1. Consider the following asynchronous iteration

$$x(t+1) = \begin{cases} \frac{1}{2}x(t), & t \in T, \\ x(t), & t \notin T, \end{cases} \tag{5.20}$$

where $x \in \mathbb{R}$, and $T = \{2^k \mid k \in \mathbb{N}_0\}$. In terms of (5.2), $f(x) = \frac{1}{2}x$. Note that f is a pseudo-contraction with $c = \frac{1}{2}$ and fixed point $x^\star = 0$. For any two consecutive elements of T, we have

$$t_{k+1} - t_k = 2^k, \quad k \in \mathbb{N}_0.$$

Thus, there is no B satisfying (5.19). However, for all $t \in \mathbb{N}$,

$$t - t_k \leq \frac{1}{2}t \leq t, \quad t \in (t_k, t_{k+1}],$$

implying that (5.3) holds with $\beta(t) = t/2$. Since the non-increasing function $\lambda(t) = 1/t$ satisfies condition (iii) of Theorem 5.1, it follows that

$$|x(t)| \leq \frac{M}{t_k}, \quad t \in (t_k, t_{k+1}].$$

One can also verify that the sequence $\{x(t)\}$ generated by (5.20) is given by

$$x(t) = \frac{x(0)/2}{t_k}, \quad t \in (t_k, t_{k+1}],$$

for all $t \geq 2$. This shows that, in this example, both the iteration (5.20) and our guaranteed upper bound have the same convergence rate. ■

Remark 5.3. As also stressed in [149], very few results on convergence rates of asynchronous iterations have appeared in the literature (see e.g., [16, 140] for exceptions). In particular, [16, §6.3.5] showed that if delays are bounded and $T^i = \mathbb{N}_0$ for each i ($t^i_{k+1} - t^i_k = 1$), then asynchronous iterations involving block-maximum norm pseudo-contractions converge linearly to the fixed point. Theorems 5.2 and 5.3 as well as Example 5.1 demonstrate that not only can Theorem 5.1 recover the results in [16], but it also quantifies the convergence rates of asynchronous iterations with unbounded upper bounds on update intervals and communication delays.

5.3 Asynchronous Algorithm for Power Control

Next, we will use our main results to analyze the convergence of asynchronous power control algorithms in wireless networks. To this end, consider a wireless network where n mobile users communicate over the same frequency band. Since concurrent transmissions interfere with each other, users must transmit with sufficient power to overcome the interference caused by the others. However, increasing the transmit power of an individual user will not only increase its own power consumption (and hence drain the battery of the device quicker), but it will also generate more interference to the other users. Thus, a natural design goal is to minimize the total power consumption while guaranteeing that all users overcome the interference caused by the others. The optimal power allocation is then the one that solves the problem:

$$\begin{aligned} \min_{p\in\mathbb{R}^n} \quad & \sum_{i=1}^{n} p_i \\ \text{subject to} \quad & p_i \geq I_i(p), \\ & \quad i = 1,\dots,n. \end{aligned} \tag{5.21}$$

Here, $p = (p_1,\dots,p_n)$, $p_i \in \mathbb{R}_+$ is the transmit power of user i, and $I_i(p)$ is the *interference function* modeling the effective interference of other users that user i must overcome. The definition of $I_i(p)$ depends on the communication technology, network configuration and user requirements; see e.g. [126] for a wide range of examples. One of the simplest interference functions is the linear one, given by

$$I_i(p) = \gamma_i \frac{\sum_{j\neq i} g_{ij}p_j + h_i}{g_{ii}}, \tag{5.22}$$

where $g_{ij} \in \mathbb{R}_+$ is the channel gain between user j and the receiver of user i, γ_i is the target Signal-to-Interference-and-Noise Ratio (SINR) of user i, and h_i is the background noise at the receiver of user i.

As observed by Yates [126], linear and several important nonlinear interference functions share common properties that allow them to be analyzed in a common framework. This observation led to the definition of *standard interference functions*.

Definition 5.2 (**Standard Interference Function**). A function $I : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ is called a standard interference function, if for all $p \geq \mathbf{0}$, the following properties are satisfied:

- *Positivity*: $I(p) > \mathbf{0}$.
- *Monotonicity*: If $p \geq p'$, then $I(p) \geq I(p')$.
- *Scalability*: For all $\alpha > 1$, $\alpha I(p) > I(\alpha p)$.

When the interference function $I(p)$ is standard and it has a fixed point, then the optimization problem (5.21) is feasible, and its unique solution is given by the fixed point of the iteration

$$p_i\big(t+1\big) = I_i\big(p(t)\big), \quad t \in \mathbb{N}_0, \tag{5.23}$$

where $i = 1, \ldots, n$ [126]. The computation of the optimal transmit power by this iteration is simpler than using traditional Lagrangian methods, since no dual variables need to be stored and manipulated. Each user is only required to update its transmit power at every time step, using information of the transmit powers used by all users in the previous iteration.

Standard interference functions do not necessarily have fixed points in the positive orthant (consider for example $I(p) = p+1$), and the existence of fixed points has to be verified separately. Furthermore, contrary to the result for contraction mappings, no guarantees on the convergence rate of the iterates involving standard interference functions are given. Already this should raise the suspicion that standard interference functions do not define contraction mappings. The following simple example establishes that this suspicion is indeed correct.

Example 5.2. Consider $I(p) = 2p + \mathbf{1}$, where $\mathbf{1}$ is the vector with all components equal to 1. One can verify that $I(p)$ is a standard interference function. However, $\|I(p) - I(p')\| = 2\|p - p'\|$, so $I(p)$ is not contractive. ■

This example provides motivations for seeking stronger conditions than standard interference function to ensure contractivity, hence linear convergence rates of the iterations. To this end, one could certainly make a separate analysis of contractivity of the particular interference functions at hand. However, if one can prove contractivity, particularly in the weighted maximum norm, then the interference function framework brings little additional value. The beauty of the framework lies in the easily verifiable conditions that guarantee synchronous and asynchronous convergence. Next, we will show that a slight reformulation of the scalability condition ensures contractivity.

We propose to study a class of interference functions which we call *contractive.*

Definition 5.3 (**Contractive Interference Function**). A function $I : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is said to be a *contractive interference function* if it, for all $p \geq \mathbf{0}$, satisfies the following conditions:

- *Positivity*: $I(p) > \mathbf{0}$.
- *Monotonicity*: If $p \geq p'$, then $I(p) \geq I(p')$.
- *Contractivity*: There exists a constant $c \in [0, 1)$ and a vector $v > \mathbf{0}$ such that for all $\varepsilon > 0$,

$$I(p + \varepsilon v) \leq I(p) + c\varepsilon v.$$

Note that the two first conditions are the same as for standard interference functions, but the scalability condition has now been replaced by contractivity.

As shown in the next theorem, contractive interference functions define contraction mappings, which implies that the associated iterations (5.23) have unique fixed points and linear convergence rates.

Theorem 5.4. *If $I : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ is a contractive interference function, then it has a unique fixed point $p^\star \in \mathbb{R}^n_+$. Moreover, for every nonnegative initial vector $p(0)$, the sequence $\{p(t)\}$ generated by (5.23) converges linearly to $p^\star$ as follows*

$$\|p(t) - p^\star\|^v_\infty \le c^t \|p(0) - p^\star\|^v_\infty.$$

Proof. Let $p \neq p'$. Since $|p_i - p'_i| \le \|p - p'\|^v_\infty v_i$ for each $i = 1, \dots, n$, we have

$$\begin{aligned} p &= p' + p - p' \\ &\le p' + \|p - p'\|^v_\infty \, v \, . \end{aligned}$$

It follows from monotonicity and contractivity properties that

$$\begin{aligned} I(p) &\le I(p' + \|p - p'\|^v_\infty \, v) \\ &\le I(p') + c\|p - p'\|^v_\infty \, v \, . \end{aligned}$$

By interchanging the roles of p and p', we have

$$I(p') \le I(p) + c\|p - p'\|^v_\infty \, v.$$

So for all components of $I(p)$, we obtain $|I_i(p) - I_i(p')| \le c\|p - p'\|^v_\infty v_i$, which implies that

$$\|I(p') - I(p)\|^v_\infty \le c\|p - p'\|^v_\infty.$$

Therefore, I is a contraction mapping with respect to the maximum norm and the result follows. ∎

To show that the concept of contractive interference functions is useful for analyzing power control algorithms, we consider the linear interference function described in (5.22). Equation (5.22) can be rewritten as

$$I_i(p) = \sum_{j=1}^{n} \overline{g}_{ij} p_j + \overline{h}_i, \quad i = 1, \dots, n,$$

where $\overline{h}_i = \gamma_i \frac{h_i}{g_{ii}}$, and

$$\overline{g}_{ij} = \begin{cases} \gamma_i \dfrac{g_{ij}}{g_{ii}}, & j \neq i, \\ 0, & j = i. \end{cases} \tag{5.24}$$

Define $\overline{G}$ as an $n \times n$ matrix that has $\overline{g}_{ij}$ as its elements. We then have the following result.

Theorem 5.5. *If $\|\overline{G}\|_\infty^v < 1$ for some $v > \mathbf{0}$, then the linear interference function* (5.22) *is a contractive interference function with $c = \|\overline{G}\|_\infty^v$.*

Proof. The linear interference function is positive and monotone. Furthermore,

$$\begin{aligned} I_i(p + \varepsilon v) &= \sum_{j=1}^{n} \overline{g}_{ij} p_j + \overline{h}_i + \varepsilon \sum_{j=1}^{n} \overline{g}_{ij} v_j \\ &\leq I_i(p) + \varepsilon \|\overline{G}\|_\infty^v v_i \,. \end{aligned}$$

Hence, it is also contractive with $c = \|\overline{G}\|_\infty^v$. ∎

Since $\overline{G}$ is a square nonnegative matrix, $\rho(\overline{G}) < 1$ is a necessary and sufficient condition for the existence of a positive vector v for which $\|\overline{G}\|_\infty^v < 1$ [16, Corollary 2.6.1]. When the matrix $\overline{G}$ is irreducible, which is often a reasonable assumption (since we are not considering totally isolated groups of links that do not interact with each other), it is worth noticing that the positive right Perron-Frobenius eigenvector v is such that $\rho(\overline{G}) = \|\overline{G}\|_\infty^v$ [16, Proposition 2.6.6]. In either case, Theorem 5.5 confirms that if $\rho(\overline{G}) < 1$, then the power control algorithm involving linear interference functions has a unique fixed point and a linear convergence rate.

In real-world networks, communication delays are inventible, and clock drift may cause some users to execute more iterations than others. When communication delays and asynchronous execution are accounted for, the power control algorithm (5.23) becomes

$$p_i(t+1) = \begin{cases} I_i\big(p_1(s_1^i(t)), \cdots, p_n(s_n^i(t))\big), & t \in T^i, \\ p_i(t), & t \notin T^i. \end{cases} \tag{5.25}$$

Since contractive interference functions are contractions with respect to the maximum norm, Theorem 5.1 allows us to quantify the convergence rate of (5.25) under different classes of communication delays and update rates. Consider, for example, a situation where all mobiles update their powers at least once during any interval of length B, and there exists a positive integer $D_{\max}$ such that

$$0 \leq \tau_j^i(t) \leq D_{\max}, \quad t \in \mathbb{N}_0. \tag{5.26}$$

The following result gives a bound on the convergence rate of (5.25) under assumptions above.

Theorem 5.6. *Assume that $I(p)$ is c-contractive. Then, the asynchronous power control algorithm* (5.25) *satisfies*

$$\frac{1}{v_i}|p_i(t) - p_i^\star| \leq M\rho^{t_k^i}, \quad t \in (t_k^i, t_{k+1}^i], \tag{5.27}$$

where M is a positive constant, t_k^i and t_{k+1}^i are two consecutive elements of T^i, and

$$\rho = c^{\frac{1}{B+D_{\max}}}. \tag{5.28}$$

Proof. The proof is similar to the one of Theorem 5.2 and thus omitted. ∎

In [126], it has been shown that the asynchronous power control algorithm (5.25) involving standard interference functions converges asymptotically to the optimal power vector even when it is executed totally asynchronously. However, the impact of the communication delay and the update interval on the convergence rate of (5.25) was missing in [126]. In contrast, this chapter develops tools that allow to quantify the convergence rate of (5.25) under various assumptions on communication delays and update rates. Specifically, Theorem 5.6 shows that for contractive interference functions, (5.25) converges linearly if the communication delays and update rates are bounded. An analogue corollary of Theorem 5.3 would demonstrate that the convergence rate of (5.25) is upper bounded by a polynomial function of time if Assumption 3 holds.

5.4 Numerical Examples

In this section, we illustrate the accuracy of our guaranteed bounds on the convergence rate of asynchronous power control algorithms with linear interference functions. We consider four mobile users that share a channel with link gain matrix $G = [g_{ij}]$ given by

$$G = \begin{bmatrix} 0.4000 & 0.0082 & 0.0419 & 0.0579 \\ 0.0160 & 0.8530 & 0.0424 & 0.0043 \\ 0.0200 & 0.0017 & 0.1405 & 0.0010 \\ 0.1030 & 0.0036 & 0.0104 & 0.4050 \end{bmatrix} \times 10^{-3}.$$

The SINR threshold and the background noise for each user is set to $\gamma_i = 3$ and $h_i = 0.04$ mWatts, respectively. Let $\overline{G} = [\overline{g}_{ij}]$ be an 4×4 matrix defined in (5.24). The spectral radius of $\overline{G}$ is $0.7146 < 1$. It follows from Theorem 5.5 that the linear interference function is contractive with $c = 0.7146$ with respect to the maximum norm $\|\cdot\|_\infty^v$, where

$$v = (0.59, 0.14, 0.38, 0.67)$$

is the right Perron-Frobenius eigenvector of $\overline{G}$.

In order to demonstrate the flexibility of our framework, assume that each user i executes the asynchronous iteration (5.25) under the assumptions that:

- $T^i = \{ik \,|k \in \mathbb{N}_0\}$;
- $\tau_i^i(t) = 0$, for all i and all $t \in \mathbb{N}_0$;
- For all i and j with $j \neq i$,

$$\tau_j^i(t) = \begin{cases} 0, & 0 \leq t \leq 4, \\ 0.5j\big(1 + (-1)^t\big) & 5 \leq t. \end{cases}$$

It is easy to verify that the time interval between any two consecutive updates executed by all nodes is upper bounded by $B = 4$, and (5.26) holds with $D_{\max} = 4$. Therefore, according to Theorem 5.6, the asynchronous algorithm (5.25) converges linearly to the unique fixed point. In particular, the transmit power of each user satisfies (5.27) with

$$\rho = (0.7146)^{\frac{1}{8}} = 0.9588.$$

Figure 5.2 gives the simulation results of the theoretical bound obtained from Theorem 5.6 and the actual convergence rate of (5.25) for users 3 and 4. Since the communication delays and update rates are time-varying and smaller than the maximum bounds, there is a gap between the theoretical and the actual decay rates that one observes in simulations.

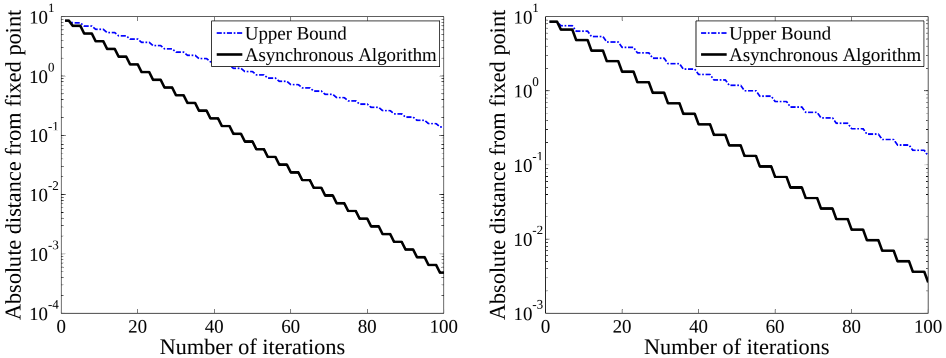


Figure 5.2: Upper bound and actual convergence rate of the asynchronous power control algorithm (5.25) for user 3 (left) and user 4 (right) under bounded communication delays. The horizontal axis represents the number of iterations and the vertical axis shows $\frac{1}{v_i}|p_i(t) - p_i^\star|$, $i = 3, 4$ (in logarithmic scale).

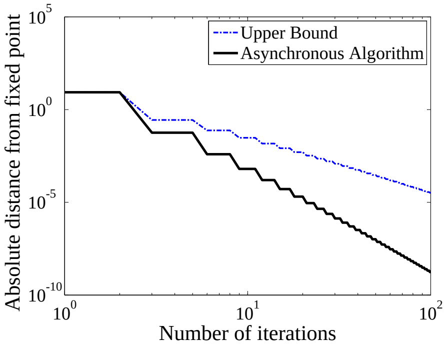
We now assume that the conditions are such that the communication delays between user 1 and the others become

$$\tau_j^1(t) = \tau_1^j(t) = \lfloor 0.1t \rfloor, \quad t \in \mathbb{N}_0,$$

for $j = 2, 3, 4$. Other delays and execution times are assumed to remain unchanged. While the above delays are unbounded, Assumption 5.3 is satisfied with $\alpha = 0.1$, so Theorem 5.3 can be used to bound the convergence rate. Using $c = 0.7146$ and $\alpha = 0.1$, the transmit power of each user satisfies (5.16) with $B = 4$ and

$$\xi = \frac{\ln 0.7146}{\ln(1 - 0.1)} = 3.189.$$

Figure 5.3 shows a comparison of the guaranteed bound obtained from Theorem 5.3 and the actual convergence rate of (5.25) for users 3 and 4.



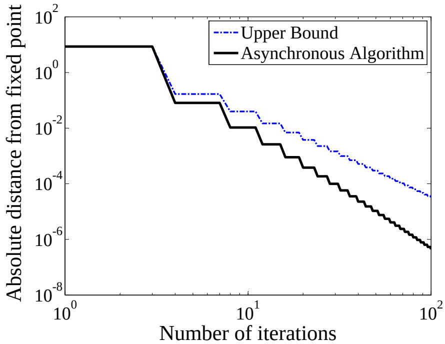


Figure 5.3: Upper bound and actual convergence rate of the asynchronous power control algorithm (5.25) for user 3 (left) and user 4 (right) under unbounded communication delays. The horizontal axis (in logarithmic scale) represents the number of iterations and the vertical axis (also in logarithmic scale) shows $\frac{1}{v_i}|p_i(t) - p_i^\star|$, $i = 3, 4$.

5.5 Summary

In this chapter, we presented convergence results for asynchronous iterations involving pseudo-contractions in the block-maximum norm. Contrary to most results in the literature, which only established asymptotic convergence or studied simplified models of asynchronism, our theorems allow to characterize the rates of convergence of asynchronous iterations and quantify how these rates depend on the update intervals and information delays. We demonstrated how our results can be used to analyze the impact of asynchrony on the convergence rate of power control algorithms in wireless networks.

Chapter 6

Asynchronous Mini-batch Algorithm for Regularized Stochastic Optimization

STOCHASTIC approximation methods such as stochastic gradient descent were among the first and the most commonly used algorithms developed for solving stochastic optimization problems [17–22]. These methods are inherently *serial* in the sense that the gradient computations take place on a single processor which has access to the whole dataset. However, it happens more and more often that one single computer is unable to store and handle the amounts of data that we encounter in practical problems. This has caused a strong interest in developing *parallel* optimization algorithms which are able to split the data and distribute the computation across multiple processors or multiple computer clusters.

One simple and popular stochastic approximation method is *mini-batching*, where iterates are updated based on the average gradient with respect to multiple data points rather than based on gradients evaluated at a single data at a time. Recently, Dekel *et. al.* [150] proposed a parallel mini-batch algorithm for regularized stochastic optimization problems, in which multiple processors compute gradients in parallel using their own local data, and then aggregate the gradients up a spanning tree to obtain the averaged gradient. While this algorithm can achieve linear speedup in the number of processors, it has the drawback that the processors need to synchronize at each round and, hence, if one of them is slower than the rest, then the entire algorithm runs at the pace of the slowest processor.

Contributions of the Chapter. In this chapter, we propose an *asynchronous* mini-batch algorithm for regularized stochastic optimization problems with smooth loss functions that eliminates the overhead associated with global synchronization. Our algorithm allows multiple processors to work at *different rates*, perform computations *independently* of each other, and update global decision variables using *out-of-date* gradients. A similar model of parallel asynchronous computation was applied to coordinate descent methods for deterministic optimization in [151–153] and mirror descent and dual averaging methods for stochastic optimization in [154]. In particular, Agarwal and Duchi [154] analyzed the convergence of asynchronous mini-batch algorithms for smooth stochastic convex problems, and interestingly showed that bounded delays do not degrade the asymptotic convergence. However, they only considered the case where the regularization term is the indicator function of a compact convex set.

We extend the results of [154] to general regularization functions (like the l_1 norm, often used to promote sparsity), and establish a sharper expected-value type of convergence rate than the one given in [154]. Specifically, we make the following contributions:

(i) For general convex regularization functions, we show that when the feasible set is closed and convex (but not necessarily bounded), the running average of the iterates generated by our algorithm with constant step-sizes converges at rate $\mathcal{O}(1/T)$ to a ball around the optimum. We derive an explicit expression that quantifies how the convergence rate and the residual error depend on loss function properties and algorithm parameters such as the step-size and the maximum delay bound $\tau_{\max}$.

(ii) For general convex regularization functions and compact feasible sets, we prove that the running average of the iterates produced by our algorithm with a time-varying step-size converges to the true optimum (without residual error) at rate

$$\mathcal{O}\left(\frac{(\tau_{\max}+1)^2}{T}+\frac{1}{\sqrt{T}}\right).$$

As long as the number of processors is $\mathcal{O}(T^{1/4})$, our algorithm enjoys near-linear speedup and converges asymptotically at a rate $\mathcal{O}(1/\sqrt{T})$.

(iii) When the regularization function is strongly convex and the feasible set is closed and convex, we establish that the iterates converge at rate

$$\mathcal{O}\left(\frac{(\tau_{\max}+1)^4}{T^2}+\frac{1}{T}\right).$$

If the number of processors is of the order of $\mathcal{O}(T^{1/4})$, this rate is $\mathcal{O}(1/T)$ asymptotically in T, which is the best known rate for strongly convex stochastic optimization problems in a serial setting [155–157].

Outline of the Chapter. In Section 6.1, we formulate the problem, discuss our assumptions, and review related work for stochastic optimization problems. The proposed asynchronous mini-batch algorithm and its main theoretical results are presented in Section 6.2. Computational experience is reported in Section 6.3 while Section 6.4 concludes the chapter with a brief statement of the results. Finally, Section 6.5 contains technical proofs of our main results.

6.1 Problem Setup

We consider stochastic convex optimization problems of the form

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \varphi(x) := \mathbb{E}_\xi\big[F(x,\xi)\big] + \Psi(x). \tag{6.1}$$

Here, x is the decision variable, ξ is a random vector whose probability distribution $\mathcal{P}$ is supported on a set $\Xi \subseteq \mathbb{R}^m$, $F(\cdot, \xi)$ is convex and differentiable for each $\xi \in \Xi$, and $\Psi(x)$ is a proper convex function that may be nonsmooth and extended real-valued. Let us define

$$f(x) := \mathbb{E}_\xi\big[F(x,\xi)\big] = \int_\Xi F(x,\xi)d\mathcal{P}(\xi). \tag{6.2}$$

The expectation function f is convex, differentiable, and $\nabla f(x) = \mathbb{E}_\xi[\nabla_x F(x,\xi)]$ [158]. Thus, $\nabla_x F(x,\xi)$ can be viewed as an unbiased estimate of $\nabla f(x)$. We use $X^\star$ to denote the set of optimal solutions of Problem (6.1) and $\varphi^\star$ to denote the corresponding optimal value.

A difficulty when solving optimization problem (6.1) is that the distribution $\mathcal{P}$ is often unknown, so the expectation (6.2) cannot be computed. This situation occurs frequently in data-driven applications such as machine learning. To support these applications, we do not assume knowledge of f (or of $\mathcal{P}$), only access to a stochastic oracle. Each time the oracle is queried with an $x \in \mathbb{R}^n$, it generates an independent and identically distributed (i.i.d.) sample ξ from $\mathcal{P}$ and returns $\nabla_x F(x,\xi)$, which is a noise-corrupted version of $\nabla f(x)$. The erroneous gradient $\nabla_x F(x,\xi)$ will be used in the update rule of our optimization algorithm instead of $\nabla f(x)$.

We also impose the following assumptions on Problem (6.1).

Assumption 6.1 (**Existence of a minimum**)**.** *The optimal set $X^\star$ is nonempty.*

Assumption 6.2 (**Lipschitz continuity of F**)**.** *For each $\xi \in \Xi$, the function $F(\cdot,\xi)$ has Lipschitz continuous gradient with constant L. That is, for all $y, z \in \mathbb{R}^n$,*

$$\|\nabla_x F(y,\xi) - \nabla_x F(z,\xi)\|_* \le L\|y - z\|. \tag{6.3}$$

Note that under Assumption 6.2, $\nabla f(x)$ is also Lipschitz continuous with the same constant L [22].

Assumption 6.3 (**Bounded gradient variance**)**.** *There exists a constant $\sigma \ge 0$ such that*

$$\mathbb{E}_\xi\big[\|\nabla_x F(x,\xi) - \nabla f(x)\|_*^2\big] \le \sigma^2, \quad \forall x \in \mathbb{R}^n.$$

Assumption 6.4 (**Closed effective domain of Ψ**)**.** *The function Ψ is simple and lower semi-continuous, and its effective domain,* $\operatorname{dom}\Psi = \{x \in \mathbb{R}^n \mid \Psi(x) < +\infty\}$, *is closed.*

Possible choices of Ψ include:

- *Unconstrained smooth minimization*: $\Psi(x) = 0$.
- *Constrained smooth minimization*: Ψ is the indicator function of a non-empty closed convex set $C \subseteq \mathbb{R}^n$, *i.e.*,

$$\Psi(x) = I_C(x) := \begin{cases} 0, & \text{if } x \in C, \\ +\infty, & \text{otherwise.} \end{cases}$$

- l_1-*regularized minimization*: $\Psi(x) = \lambda\|x\|_1$ with $\lambda > 0$.
- *Constrained* l_1-*regularized minimization*: In this case, $\Psi(x) = \lambda\|x\|_1 + I_C(x)$ with $\lambda > 0$.

Several practical problems in machine learning, statistical applications, and signal processing satisfy Assumptions 6.1–6.4 [37]. One such example is l_1-*regularized logistic regression* for sparse binary classification. We are then given a large number of observations

$$\left\{\xi_j = \left(\xi_j^{(1)}, \xi_j^{(2)}\right) \mid \xi_j^{(1)} \in \mathbb{R}^n,\ \xi_j^{(2)} \in \{-1,+1\},\ j = 1,\dots,J\right\},$$

drawn i.i.d. from an unknown distribution $\mathcal{P}$, and want to solve the minimization problem (6.1) with

$$F(x,\xi) = \log\left(1 + \exp\left(-\xi_j^{(2)}\langle\xi_j^{(1)}, x\rangle\right)\right),$$

and $\Psi(x) = \lambda\|x\|_1$. The role of l_1 regularization is to produce sparse solutions.

6.1.1 Related Work

We now review several existing first-order methods for solving regularized stochastic optimization problems.

Serial stochastic gradient methods: Stochastic gradient methods have become extremely popular for large-scale optimization problems such as boosting and training linear Support Vector Machines (SVMs) [92]. Their popularity comes mainly from the fact that they are easy to implement and have low computational cost per iteration. In the classical stochastic gradient method, a single processor iteratively updates the current vector $x(k)$ by sampling ξ from $\mathcal{P}$, computing $g(k) = \nabla_x F(x(k),\xi)$, and performing the update

$$x\big(k+1\big) = \Pi_C\big(x(k) - \gamma(k)g(k)\big). \tag{6.4}$$

Here, Π_C is the Euclidean projection operator onto the set C, and $\gamma(k)$ is a positive step-size. The update rules of more general classes of stochastic gradient methods such as the *mirror descent* and *dual-averaging* can be described by

$$x\big(k+1\big) = \Omega\big(g(k), h(k), \gamma(k)\big), \tag{6.5}$$

where Ω takes three arguments: the stochastic gradient $g(k)$, $h(k)$ that summarizes the information about the past, and the step-size $\gamma(k)$. The stochastic gradient projection method (6.4) fits this template by defining $h(k) = x(k)$ and

$$\Omega\big(g(k), h(k), \gamma(k)\big) = \underset{z\in C}{\operatorname{argmin}}\ \|z - (x(k) - \gamma(k)g(k))\|_2^2.$$

Recently, Lan [19] proposed two stochastic gradient algorithms for regularized stochastic optimization, which he named the *modified mirror descent method* and the *accelerated stochastic approximation method*. These algorithms achieve convergence rates of the form

$$\mathcal{O}\left(\frac{L}{T} + \frac{\sigma}{\sqrt{T}}\right) \text{ and } \mathcal{O}\left(\frac{L}{T^2} + \frac{\sigma}{\sqrt{T}}\right),$$

respectively, which are $\mathcal{O}(\sigma/\sqrt{T})$ asymptotically in T. Here, T is the number of iterations. The algorithms presented in [19] use step-sizes based on an a priori knowledge of the performed number of iterations T and assume that the feasible set is bounded. A different stochastic gradient method was developed in [20], where the step-sizes in its update rule are independent of the number of iterations. Under the assumption of bounded feasible sets, the method was shown to have the same convergence rate as that of [19].

Strong convexity of the objective function can speedup the convergence rate significantly. In particular, [21] proposed a stochastic gradient method for minimizing strongly convex composite objective functions, which asymptotically converges at a rate of $\mathcal{O}(\sigma^2/T)$. Note that the rate $\mathcal{O}(\sigma/\sqrt{T})$ for convex stochastic problems and the rate $\mathcal{O}(\sigma^2/T)$ for strongly convex stochastic problems are the best known asymptotic convergence rates for stochastic gradient methods [155–157].

The update rules of methods in [19–21] use the stochastic gradient of the current iteration ($g(k)$). Nesterov [159] proposed a different class of update rules, called dual-averaging, for nonsmooth stochastic optimization problems. In these methods, the current iterate $x(k)$ is updated using a weighted average of all past stochastic gradients. In its simplest form, the dual averaging performs the update

$$x(k+1) = \Pi_C\left(-\gamma(k)\sum_{t=0}^{k} g(t)\right),$$

which fits the template (6.5) with $h(k) = \{g(0), \dots, g(k-1)\}$. Xiao [22] developed an extension of the dual averaging method to solve regularized optimization problems of the form (6.1). Under Assumptions 6.1–6.4, the asymptotic convergence rate obtained in [22] is $\mathcal{O}(\sigma/\sqrt{T})$ for convex objective functions and $\mathcal{O}(\sigma^2 \ln T/T)$ for strongly convex objective functions. In [160], the convergence rate of dual-averaging methods under strong convexity assumption was improved to $\mathcal{O}(\sigma^2/T)$.

Stochastic gradient methods with mini-batching: In many emerging applications, such as large-scale machine learning and statistics, the size of dataset is so huge that it cannot fit on one computer. Hence, we need optimization algorithms that can be conveniently and efficiently executed in parallel on multiple processors. One key disadvantage of stochastic gradient methods mentioned above is that they are inherently sequential, *i.e*, updating each iteration requires having completed the previous iteration. This makes it difficult to parallelize these methods under advanced architectures such as GPUs, multi-core CPUs, or distributed clusters [161].

The second weakness of serial stochastic gradient methods is that they usually have slow asymptotic convergence due to the inherent variance in the stochastic gradients.

A common practical solution for parallelizing stochastic gradient methods and reducing the stochastic variance is to employ mini-batches [162]. In mini-batch algorithms, multiple processors compute the stochastic gradients

$$\nabla_x F\big(x(k), \xi_1\big), \dots, \nabla_x F\big(x(k), \xi_b\big)$$

on b samples $\xi_1, \dots, \xi_b$ in parallel, communicate with each other to obtain the averaged stochastic gradient vector

$$g_{\text{ave}}(k) = \frac{1}{b} \sum_{i=1}^{b} \nabla_x F(x(k), \xi_i),$$

and then update $x(k)$ via

$$x\big(k+1\big) = \Omega\big(g_{\text{ave}}(k), h(k), \gamma(k)\big). \tag{6.6}$$

In comparison with (6.5), the update rule (6.6) shows that a mini-batch algorithm updates x using the average gradient over multiple calls to the stochastic oracle, rather than updating after each query of the oracle. For a batch size b, using mini-batches can reduce the variance of stochastic gradient from σ to $\sigma/\sqrt{b}$. Dekel *et. al.* [150] exploited this fact to develop a parallel mini-batch algorithm for solving (6.1) that achieves linear speedup in the number of processors. The asymptotic convergence rate was proven to be $\mathcal{O}\big(\sigma^2/(bT)\big)$ for strongly convex problems and $\mathcal{O}\big(\sigma/\sqrt{bT}\big)$ for convex problems.

Asynchronous stochastic gradient methods: A key property of synchronous mini-batch methods is that at k^{th} iteration, all the gradients involved in the update must be computed at the same vector $x(k)$ by the processors. This can cause some processors to be idle during each iteration since they may perform computations faster than others. Note also that at each round of synchronous mini-batch methods, averaging the gradients requires message passing between the processors. In order to reduce the waste generated by the need for global synchronization and requiring massive communication overhead, we can allow the processors to operate in an *asynchronous* manner.

There have been extensive studies on asynchronous stochastic optimization, but mostly without the *smoothness* assumption (6.3), see, *e.g.*, [163–165]. More precisely, these works studied the convergence of asynchronous algorithms under the assumption that the objective function F is *nonsmooth* with *bounded* subgradients, *i.e.*, $\mathbb{E}[\|\partial_x F(x, \xi)\|_*^2] \leq G^2$ holds for some $G > 0$ and for all x. The literature on asynchronous algorithms for smooth stochastic optimization is relatively sparse. Langford *et. al.* [166] presented an asynchronous parallel algorithm for smooth optimization problems of the form (6.1), with $\Psi(x) = 0$. In the algorithm, the processors compute gradients with a fixed delay $\tau_{\max}$ and update x in a round-robin fashion via

$$x\big(k+1\big) = \Omega\big(g(k - \tau_{\max}), h(k), \gamma(k)\big).$$

It was shown that the constant delay $\tau_{\max}$ introduces negligible penalty in the convergence rate. Agarwal and Duchi [154] extended this work to constrained smooth minimization in which F is a smooth function with bounded gradients and Ψ is the indicator function of a compact convex set $C \subseteq \mathbb{R}^n$. They studied the convergence of asynchronous mini-batch methods with the update rule

$$x\big(k+1\big) = \Omega\big(g_{\text{ave}}(k-\tau(k)), h(k), \gamma(k)\big),$$

where $\tau(k)$ is a bounded time-varying delay. The asymptotic convergence rate was proven to be $\mathcal{O}\big(\sigma/\sqrt{bT}\big)$ for convex objective functions. This result shows that even in the face of asynchronism, convergence guarantees similar to synchronous mini-batch methods are possible. Convergence rates for strongly convex functions was not discussed in [154].

Our goal is to (i) develop an algorithm for solving regularized stochastic optimization problems which combines the strong performance guarantees of serial stochastic gradient methods, the parallelization benefits of mini-batching algorithms, and the speed-ups enabled by asynchronous implementations; to (ii) extend the analysis in [154] to solve the optimization problem (6.1) with *general* regularization functions (not necessarily $\Psi(x) = I_C(x)$) without any additional assumption on boundedness of either the gradients or the feasible sets; and to (iii) determine whether an asynchronous mini-batch algorithm achieves the optimal rate $\mathcal{O}\big(\sigma^2/(bT)\big)$ under the strong convexity assumption.

6.2 An Asynchronous Mini-batch Algorithm

In this section, we present an *asynchronous* mini-batch algorithm that exploits multiple processors to solve Problem (6.1). We characterize the iteration complexity and the convergence rate of the proposed algorithm, and show that these compare favourably with the state of the art. Our approach is distinguished from recent work on stochastic optimization [19–22, 150, 154] in that it can deal with *asynchrony* and *smooth* objective functions as well as *general* regularization functions at the same time. To the best of our knowledge, our asynchronous algorithm is the first to attain the optimal convergence rates for convex and strongly convex stochastic composite optimization in spite of time-varying delays.

6.2.1 Description of Algorithm

We assume p processors have access to a shared memory for the decision variable x. The processors may have different capabilities (in terms of processing power and access to data) and are able to update x without the need for coordination or synchronization. Conceptually, the algorithm lets each processor run its own stochastic composite mirror descent process, repeating the following steps:

1. Read x from the shared memory and load it into the local storage location $\widehat{x}$;

2. Sample b i.i.d random variables $\xi_1, \ldots, \xi_b$ from the distribution $\mathcal{P}$;

3. Compute the averaged stochastic gradient vector

$$\widehat{g}_{\text{ave}} = \frac{1}{b} \sum_{i=1}^{b} \nabla_x F(\widehat{x}, \xi_i);$$

4. Update current x in the shared memory via

$$x \leftarrow \operatorname*{argmin}_z \left\{ \left\langle \widehat{g}_{\text{ave}}, z \right\rangle + \Psi(z) + \frac{1}{\gamma} D_\omega(x, z) \right\}.$$

The algorithm can be implemented in many ways as depicted in Figure 6.1. One way is to consider the p processors as peers that each execute the four-step algorithm independently of each other and only share the global memory for storing x. In this case, each processor reads the decision vector twice in each round: once in the first step (before evaluating the averaged gradient), and once in the last step (before carrying out the minimization). To ensure correctness, Step 4 must be an atomic operation, where the executing processor puts a write lock on the global memory until it has written back the result of the minimization (cf. Figure 6.1, left). The algorithm can also be executed in a master-worker setting. In this case, each of the worker nodes retrieves x from the master in Step 1 and returns the averaged gradient to the master in Step 3; the fourth step (carrying out the minimization) is executed by the master (cf. Figure 6.1, right).

Independently of how we choose to implement the algorithm, processors may work at different rates: while one processor updates the decision vector (in the shared memory setting) or send its averaged gradient to the master (in the master-worker setting), the others are generally busy computing averaged gradient vectors. The processors that perform gradient evaluations do not need to be aware of updates to the decision vector, but can continue to operate on stale information about x. Therefore, unlike *synchronous* parallel mini-batch algorithms [150], there is no need for processors to wait for each other to finish the gradient computations. Moreover, the value $\widehat{x}$ at which the average of gradients is evaluated by a processor may differ from the value of x to which the update is applied.

Algorithm 1 describes the p asynchronous processes that run in parallel. To describe the progress of the overall optimization process, we introduce a counter k that is incremented each time x is updated. We let $d(k)$ denote the time at which $\widehat{x}$ used to compute the averaged gradient involved in the update of $x(k)$ was read from the shared memory. It is clear that $0 \le d(k) \le k$ for all $k \in \mathbb{N}_0$. The value

$$\tau(k) := k - d(k)$$

can be viewed as the delay between reading and updating for processors and captures the staleness of the information used to compute the average of gradients for the k^{th} update. We assume that the time-varying delay $\tau(k)$ is bounded; this is stated in the following assumption.

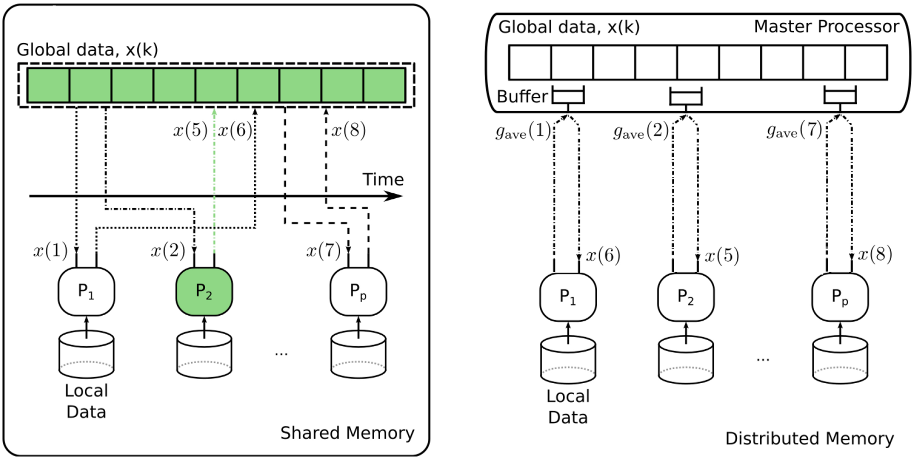


Figure 6.1: Illustration of two conceptually different realizations of Algorithm 1: (1) a shared memory implementation (left); (2) a master-worker implementation (right). In the shared memory setting shown to the left, processor P_2 reads $x(2)$ from the shared memory and computes the averaged gradient vector $g_{\text{ave}}(2) = \frac{1}{b}\sum_{i=1}^{b} \nabla_x F\big(x(2), \xi_i\big)$. As the processors are being run without synchronization, $x(3)$ and $x(4)$ are written to the shared memory by other processors while P_2 is evaluating $g_{\text{ave}}(2)$. The figure shows a snapshot of the algorithm at time instance $k = 5$, at which the shared memory is locked by P_2 to read the current x, *i.e.* $x(4)$, to update it using the out-of-date gradient $g_{\text{ave}}(2)$, and write $x(5)$ to the memory. In the master-worker setting illustrated to the right, workers evaluate averaged gradient vectors in parallel and send their computations to buffers on the master processor, which is the sole entity with access to the global memory. The master performs an update using (possibly) out-of-date gradients and passes the updated decision vector x back to the workers.

Assumption 6.5 (**Bounded Delay**). *There is a nonnegative integer $\tau_{\max}$ such that*

$$0 \le \tau(k) \le \tau_{\max}, \quad k \in \mathbb{N}_0.$$

The value of $\tau_{\max}$ is an indicator of the asynchronism in the algorithm and in the execution platform. In practice, $\tau_{\max}$ will depend on the number of parallel processors used in the algorithm [151–153]. Note that the cyclic-delay mini-batch algorithm [154], in which the processors are ordered and each updates the decision variable under a fixed schedule, is a special case of Algorithm 1 where $d(k) = k - p + 1$, or, equivalently, $\tau(k) = p - 1$ for all k.

6.2.2 Convergence Rate for General Convex Regularization

The following theorem establishes convergence properties of Algorithm 1 when a constant step-size is used.

Algorithm 1 Asynchronous Mini-batch Algorithm (running on each processor)

1: **Inputs**: positive step-sizes $\{\gamma(k)\}_{k\in\mathbb{N}_0}$; batch size $b \in \mathbb{N}$.
2: **Initialization:** $x(0) \in \operatorname{dom}\Psi$; $k = 0$.
3: **repeat**
4: receive inputs $\xi_1, \ldots, \xi_b$ sampled i.i.d. from distribution $\mathcal{P}$;

$$g_{\text{ave}}\big(d(k)\big) \leftarrow \frac{1}{b}\sum_{i=1}^{b} \nabla_x F\big(x(d(k)), \xi_i\big);$$

$$x(k+1) \leftarrow \operatorname*{argmin}_z \left\{ \big\langle g_{\text{ave}}\big(d(k)\big), z\big\rangle + \Psi(z) + \frac{1}{\gamma(k)} D_\omega(x(k), z) \right\} \tag{6.7}$$

$$k \leftarrow k+1;$$

5: **until** termination test satisfied

Theorem 6.1. *Let Assumptions 6.1–6.5 hold. Assume also that for all $k \in \mathbb{N}_0$,*

$$\gamma(k) = \gamma \in \left(0, \ \frac{1}{L(\tau_{\max}+1)^2}\right). \tag{6.8}$$

Then, for every $T \in \mathbb{N}$ and any optimizer $x^\star$ of (6.1), we have

$$\mathbb{E}\big[\varphi\big(x_{\text{ave}}(T)\big)\big] - \varphi^\star \leq \frac{D_\omega\big(x(0), x^\star\big)}{\gamma T} + \frac{\gamma c \sigma^2}{2b\big(1 - \gamma L(\tau_{\max}+1)^2\big)},$$

where $x_{\text{ave}}(T)$ is the Cesáro average of the iterates, i.e.,

$$x_{\text{ave}}(T) := \frac{1}{T}\sum_{k=1}^{T} x(k).$$

Furthermore, b is the batch size, the expectation is taken with respect to all random variables $\{\xi_i(k) \mid i = 1, \ldots, b, \ k = 0, \ldots, T-1\}$, and $c \in [1, b]$ is given by

$$c = \begin{cases} 1, & \text{if } \|\cdot\|_* = \|\cdot\|_2, \\ 2\max_{\|x\|\leq 1}\omega(x), & \text{otherwise.} \end{cases}$$

Proof. See Appendix 6.5.1. ∎

Theorem 6.1 demonstrates that for any constant step-size γ satisfying (6.8), the running average of iterates generated by Algorithm 1 will converge in expectation to a ball around the optimum at a rate of $\mathcal{O}(1/T)$. The convergence rate and the residual error depend on the choice of γ: decreasing γ reduces the residual error, but it also results in a slower convergence. We now describe a possible strategy for selecting the constant step-size. Let T_ε be the total number of iterations necessary to achieve ε-optimal solution to Problem (6.1), that is, $\mathbb{E}\big[\varphi\big(x_{\text{ave}}(T)\big)\big] - \varphi^\star \leq \varepsilon$ when $T \geq T_\varepsilon$. If we pick

$$\gamma = \frac{\varepsilon}{L\varepsilon(\tau_{\max}+1)^2 + c\sigma^2/b}, \tag{6.9}$$

then, using Theorem 6.1, the corresponding $x_{\text{ave}}(T)$ satisfies

$$\mathbb{E}\big[\varphi\big(x_{\text{ave}}(T)\big)\big] - \varphi^\star \leq \frac{\varepsilon_0}{T}\left(L(\tau_{\max}+1)^2 + \frac{c\sigma^2}{b\varepsilon}\right) + \frac{\varepsilon}{2},$$

where $\varepsilon_0 = D_\omega\big(x(0), x^\star\big)$. This inequality tells us that if the first term on the right-hand side is less than $\varepsilon/2$, *i.e.*, if

$$T \geq T_\varepsilon := 2\varepsilon_0\left(\frac{L(\tau_{\max}+1)^2}{\varepsilon} + \frac{c\sigma^2}{b\varepsilon^2}\right),$$

then $\mathbb{E}\big[\varphi\big(x_{\text{ave}}(T)\big)\big] - \varphi^\star \leq \varepsilon$. Hence, the iteration complexity of Algorithm 1 with the step-size choice (6.9) is given by

$$\mathcal{O}\left(\frac{L(\tau_{\max}+1)^2}{\varepsilon} + \frac{c\sigma^2}{b\varepsilon^2}\right). \tag{6.10}$$

As long as the maximum delay bound $\tau_{\max}$ is of the order $1/\sqrt{\varepsilon}$, the first term in (6.10) is asymptotically negligible. In this case, the iteration complexity of Algorithm 1 is asymptotically $\mathcal{O}(c\sigma^2/b\varepsilon^2)$, which is exactly the iteration complexity achieved by a serial mini-batch algorithm [150]. As discussed before, $\tau_{\max}$ is related to the number of processors. Therefore, if the number of processors is of the order of $\mathcal{O}(1/\sqrt{\varepsilon})$, parallelization does not appreciably degrade asymptotic convergence of Algorithm 1. Furthermore, as p processors are being run asynchronously and in parallel, updates may occur roughly p times as quickly, which means that the near-linear speedup in the number of processors can be expected.

Remark 6.1. Another strategy for the selection of the constant step-size in Algorithm 1 is to use γ that depends on the prior knowledge of the number of iterations to be performed. More precisely, assume that the number of iterations is fixed in advance, say equal to T_F. By choosing γ as

$$\gamma = \frac{1}{L(\tau_{\max}+1)^2 + \alpha\sqrt{T_F}},$$

for some $\alpha > 0$, it follows from Theorem 6.1 that the running average of the iterates after T_F iterations satisfies

$$\mathbb{E}\big[\varphi\big(x_{\text{ave}}(T_F)\big)\big] - \varphi^\star \leq \frac{L(\tau_{\max}+1)^2 D_\omega\big(x(0), x^\star\big)}{T_F} + \frac{1}{\sqrt{T_F}}\left(\alpha D_\omega\big(x(0), x^\star\big) + \frac{c\sigma^2}{2\alpha b}\right).$$

It is easy to verify that the optimal choice of α, which minimizes the second term on the right-hand-side of the above inequality, is

$$\alpha^\star = \frac{\sigma\sqrt{c}}{\sqrt{2bD_\omega\big(x(0), x^\star\big)}}.$$

With this choice of α, we then have

$$\mathbb{E}\big[\varphi\big(x_{\mathrm{ave}}(T_F)\big)\big] - \varphi^\star \leq \frac{L(\tau_{\max}+1)^2 D_\omega\big(x(0), x^\star\big)}{T_F} + \frac{\sigma\sqrt{2cD_\omega\big(x(0), x^\star\big)}}{\sqrt{bT_F}}.$$

In the case that $\tau_{\max} = 0$, the preceding guaranteed bound reduces to the one obtained in [19, Theorem 1] for the serial stochastic mirror descent algorithm with constant step-sizes. Note that in order to implement Algorithm 1 with the optimal constant step-size policy, we need to estimate an upper bound on $D_\omega\big(x(0), x^\star\big)$, since $D_\omega\big(x(0), x^\star\big)$ is usually unknown.

The following theorem characterizes the convergence of Algorithm 1 with a time-varying step-size sequence when dom Ψ is *bounded* in addition to being closed and convex.

Theorem 6.2. *Suppose that Assumptions 6.1–6.5 hold. In addition, suppose that* dom Ψ *is compact and that* $D_\omega(\cdot,\cdot)$ *is bounded on* dom Ψ. *Let*

$$R^2 = \max_{x,y\in \mathrm{dom}\ \Psi} D_\omega(x,y).$$

If $\{\gamma(k)\}_{k\in\mathbb{N}_0}$ *is set to* $\gamma(k)^{-1} = L(\tau_{\max}+1)^2 + \alpha(k)$ *with*

$$\alpha(k) = \frac{\sigma\sqrt{c}\sqrt{k+1}}{R\sqrt{b}},$$

then the Cesáro average of the iterates generated by Algorithm 1 satisfies

$$\mathbb{E}\big[\varphi\big(x_{\mathrm{ave}}(T)\big)\big] - \varphi^\star \leq \frac{LR^2(\tau_{\max}+1)^2}{T} + \frac{2\sigma R\sqrt{c}}{\sqrt{bT}},$$

for all $T \in \mathbb{N}$.

Proof. See Appendix 6.5.2. ∎

The time-varying step-size $\gamma(k)$, which ensures the convergence of the algorithm, consists of two terms: the time-varying term $\alpha(k)$ should control the errors from stochastic gradient information while the role of the constant term $L(\tau_{\max}+1)^2$ is to decrease the effects of asynchrony (bounded delays) on the convergence of the algorithm. According to Theorem 6.2, in the case that $\tau_{\max} = \mathcal{O}(T^{1/4})$, the delay becomes increasingly harmless as the algorithm progresses and the expected function value evaluated at $x_{\mathrm{ave}}(T)$ converges asymptotically at a rate $\mathcal{O}(1/\sqrt{T})$, which is known to be the best achievable rate of the mirror descent method for nonsmooth stochastic convex optimization problems [18].

For the special case of the optimization problem (6.1) where Ψ is restricted to be the indicator function of a compact convex set, Agarwal and Duchi [154, Theorem 2] showed that the convergence rate of the delayed stochastic mirror descent method with time-varying step-sizes is

$$\mathcal{O}\left(\frac{LR^2 + RG\tau_{\max}}{T} + \frac{\sigma R\sqrt{c}}{\sqrt{Tb}} + \frac{LR^2G^2\tau_{\max}^2 b\log T}{c\sigma^2 T}\right),$$

where G is the maximum bound on $\sqrt{\mathbb{E}[\|\nabla_x F(x,\xi)\|_*^2]}$. Comparing with this result, instead of a asymptotic penalty of the form $\mathcal{O}(\tau_{\max}^2 \log T/T)$ due to the delays, we have the penalty $\mathcal{O}(\tau_{\max}^2/T)$, which is much smaller for large T. Therefore, not only do we extend the result of [154] to general regularization functions, but we also obtain a sharper guaranteed convergence rate than the one presented in [154].

6.2.3 Convergence Rate for Strongly Convex Regularization

In this subsection, we restrict our attention to stochastic composite optimization problems with strongly convex regularization terms. Specifically, we assume that Ψ is μ_Ψ-strongly convex with respect to $\|\cdot\|$, that is, for any $x, y \in \mathrm{dom}\,\Psi$,

$$\Psi(y) \geq \Psi(x) + \langle s, y - x\rangle + \frac{\mu_\Psi}{2}\|y - x\|^2, \quad \forall s \in \partial\Psi(x).$$

The strong convexity of Ψ implies that Problem (6.1) has a unique minimizer $x^\star$ [167, Corollary 11.16]. Examples of the strongly convex function Ψ include:

- *l_2-regularization*: $\Psi(x) = (\lambda/2)\|x\|_2^2$ with $\lambda > 0$.
- *Elastic net regularization*: $\Psi(x) = \lambda_1\|x\|_1 + (\lambda_2/2)\|x\|_2^2$ with $\lambda_1 > 0$ and $\lambda_2 > 0$.

In order to derive the convergence rate of Algorithm 1 for solving (6.1) with a strongly convex regularization term, we need to assume that the Bregman distance function $D(x,y)$ used in the algorithm satisfies the next assumption.

Assumption 6.6 (**Quadratic growth condition**). *For all $x, y \in \mathrm{dom}\,\Psi$, we have*

$$D_\omega(x,y) \leq \frac{Q}{2}\|x - y\|^2,$$

with $Q \geq \mu_\omega$.

For example, if $\omega(x) = \frac{1}{2}\|x\|_2^2$, then $D_\omega(x,y) = \frac{1}{2}\|x-y\|_2^2$ and $Q = 1$. Note that Assumption 6.6 will automatically hold when the distance generating function ω has Lipschitz continuous gradient with a constant Q [157].

The associated convergence result now reads as follows.

Theorem 6.3. *Suppose that the regularization function Ψ is μ_Ψ-strongly convex and that Assumptions 6.2–6.6 hold. If $\{\gamma(k)\}_{k\in\mathbb{N}_0}$ is set to $\gamma(k)^{-1} = 2L(\tau_{\max}+1)^2 + \beta(k)$ with*

$$\beta(k) = \frac{\mu_\Psi}{3Q}\big(k + \tau_{\max} + 1\big),$$

then the iterates produced by Algorithm 1 satisfies

$$\mathbb{E}\big[\|x(T) - x^\star\|^2\big] \leq \frac{2\left(\frac{6LQ}{\mu_\Psi}+1\right)^2(\tau_{\max}+1)^4}{(T+1)^2} D_\omega\big(x(0), x^\star\big) + \frac{18c\sigma^2Q^2}{b\mu_\Psi^2(T+1)},$$

for all $T \in \mathbb{N}$.

Proof. See Appendix 6.5.3. ∎

An interesting point regarding Theorem 6.3 is that the maximum delay bound $\tau_{\max}$ can be as large as $\mathcal{O}(T^{1/4})$ without affecting the asymptotic convergence rate of Algorithm 1. In this case, our asynchronous mini-batch algorithm converges asymptotically at a rate of $\mathcal{O}(1/T)$, which matches the best known rate achievable in a serial setting.

6.3 Experimental Results

We have developed a complete master-worker implementation of our algorithm in C++ using the Massage Passing Interface (MPI) libraries OpenMPI [168]. Although we argued in Section 6.2 that Algorithm 1 can be implemented using atomic operations on shared-memory computing architectures, we have chosen the MPI implementation due to its flexibility in scaling the problem to distributed-memory environments.

We provide empirical results to show how Algorithm 1 performs on stochastic optimization problems. To this end, we use the text categorization dataset `rcv1` [169] which consists of $J \approx 800000$ documents with $n \approx 50000$ unique stemmed tokens spanning 103 topics. Out of these topics, we decide to sort out sports-related documents. We apply our code for Algorithm 1 to the following l_1-regularized logistic regression problem on the dataset:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \qquad \mathbb{E}_{\left(\xi_j^{(1)},\xi_j^{(2)}\right)}\left[\log\left(1+\exp\left(-\xi_j^{(2)}\langle\xi_j^{(1)}, x\rangle\right)\right)\right] + \lambda\|x\|_1 + I_C(x).$$

Here, $\xi_j^{(1)} \in \mathbb{R}^n$, $j = 1, \dots, J$, represents a feature vector and $\xi_j^{(2)} \in \{-1, 1\}$ denotes its associated label, λ is the regularization parameter, and

$$C = \{x \in \mathbb{R}^n : \|x\|_2 \leq R\}.$$

The label $\xi_j^{(2)}$ indicates whether a selected sample falls into the desired category, or not. In particular, $\xi_j^{(2)} = 1$ if the sampled document is about sports, and $\xi_j^{(2)} = -1$ otherwise. We use the distance generating function $\omega(x) = \frac{1}{2}\|x\|_2^2$ in all experiments.

To evaluate the performance of our asynchronous algorithm, we set $\lambda = 0.01$ and $R = 100$. As the feasible set is bounded, we implement Algorithm 1 with the time-varying step-size given in Theorem 6.2. We use a batch size of $b = 1000$ samples. For relative speedup comparison purposes, we run the algorithm on $p = 1, 2, 4, 6, 8, 10$ workers until a fixed tolerance $\varepsilon = 0.01$ is met.

Figure 6.2 presents the achieved relative speedup of the algorithm with respect to the number of workers used. The relative speedup of the algorithm on p workers is defined as $S(p) = t_1/t_p$, where t_1 and t_p are the time it takes to run the corresponding algorithm (to ε-accuracy) on 1 and p workers, respectively. We observe a near-linear relative speedup, consistent with our theoretical results. However, as the number of workers increases, the relative speedup starts saturating due to the communication overhead at the master side. Speedup values are averaged over 10 Monte Carlo simulations.

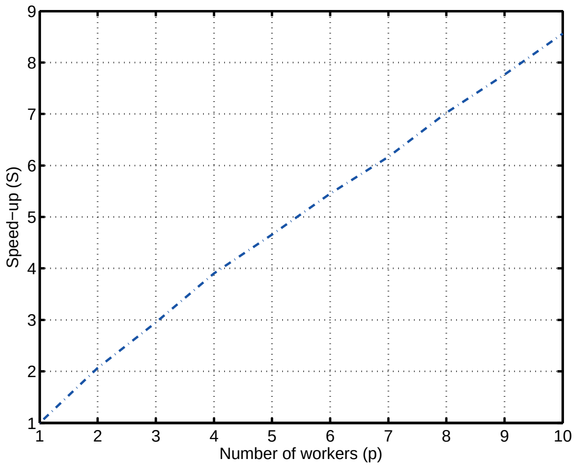


Figure 6.2: Speedup of Algorithm 1 with respect to the number of workers.

6.4 Summary

We proposed an asynchronous mini-batch algorithm that exploits multiple processors to solve regularized stochastic optimization problems with smooth loss functions. We established that for closed and convex constraint sets, the iteration complexity of the algorithm with constant step-sizes is asymptotically $\mathcal{O}(1/\varepsilon^2)$. For compact constraint sets, we proved that the running average of the iterates generated by our algorithm with time-varying step-size converges to the optimum at a rate $\mathcal{O}(1/\sqrt{T})$. When the regularization function is strongly convex and the constraint set is closed and convex, the algorithm achieves the rate $\mathcal{O}(1/T)$. We showed that the penalty in convergence rate of the algorithm due to asynchrony is asymptotically negligible and a near-linear speedup in the number of processors can be expected. Our computational experience confirmed the theory.

6.5 Appendix

In this section, we prove the main results of the paper, namely, Theorems 6.1–6.3. We first state three key lemmas which are instrumental in our argument.

The following result establishes an important recursion for the iterates generated by Algorithm 1.

Lemma 6.1. *Suppose Assumptions 6.1–6.5 hold. Then, the iterates* $\{x(k)\}_{k\in\mathbb{N}_0}$ *generated by Algorithm 1 satisfy*

$$\begin{aligned}\varphi\big(x(k+1)\big)-\varphi^\star+\frac{1}{\gamma(k)}D_\omega\big(x(k+1),x^\star\big)\le{}&\frac{1}{2\eta(k)}\big\|e\big(d(k)\big)\big\|_*^2\\&+\big\langle e\big(d(k)\big),x(k)-x^\star\big\rangle+\frac{1}{\gamma(k)}D_\omega\big(x(k),x^\star\big)\\&+\frac{L(\tau_{\max}+1)}{2}\sum_{j=0}^{\tau_{\max}}\big\|x(k-j)-x(k-j+1)\big\|^2\\&-\frac{1}{2}\left(\frac{1}{\gamma(k)}-\eta(k)\right)\|x(k+1)-x(k)\|^2\\&-\frac{\mu_\Psi}{2}\big\|x(k+1)-x^\star\big\|^2,\end{aligned}\tag{6.11}$$

where $x^\star\in X^\star$, $e(k):=\nabla f(x(k))-g_{\text{ave}}(k)$ *is the error in the gradient estimate, and* $\{\eta(k)\}$ *is a sequence of strictly positive numbers.*

Proof. We start with the first-order optimality condition for the point $x(k+1)$ in the minimization problem (6.7): there exists subgradient $s(k+1)\in\partial\Psi(x(k+1))$ such that for all $z\in\operatorname{dom}\Psi$, we have

$$\left\langle g_{\text{ave}}\big(d(k)\big)+s(k+1)+\frac{1}{\gamma(k)}\nabla_{(2)}D_\omega\big(x(k),x(k+1)\big),z-x(k+1)\right\rangle\ge 0,$$

where $\nabla_{(2)}D_\omega(\cdot,\cdot)$ denotes the partial derivative of the Bregman distance function with respect to the second variable. Plugging the following equality

$$\nabla_{(2)}D_\omega\big(x(k),x(k+1)\big)=\nabla\omega\big(x(k+1)\big)-\nabla\omega\big(x(k)\big),$$

into the previous inequality and re-arranging terms gives

$$\begin{aligned}\frac{1}{\gamma(k)}\left\langle\nabla\omega\big(x(k)\big)-\nabla\omega\big(x(k+1)\big),z-x(k+1)\right\rangle&\le\left\langle g_{\text{ave}}\big(d(k)\big)+s(k+1),z-x(k+1)\right\rangle\\&=\left\langle g_{\text{ave}}\big(d(k)\big),z-x(k+1)\right\rangle\\&\quad+\left\langle s(k+1),z-x(k+1)\right\rangle\\&\le\left\langle g_{\text{ave}}\big(d(k)\big),z-x(k+1)\right\rangle\\&\quad+\Psi(z)-\Psi(x(k+1))-\frac{\mu_\Psi}{2}\big\|z-x(k+1)\big\|^2,\end{aligned}\tag{6.12}$$

where the last inequality used

$$\Psi(z) \geq \Psi\big(x(k+1)\big) + \big\langle s(k+1), z - x(k+1)\big\rangle + \frac{\mu_\Psi}{2}\big\|z - x(k+1)\big\|^2,$$

by the (strong) convexity of Ψ. We now use the following well-known *three point identity* of the Bregman distance function [170] to rewrite the left-hand side of (6.12):

$$\big\langle \nabla\omega(a) - \nabla\omega(b), c - b\big\rangle = D_\omega(a,b) - D_\omega(a,c) + D_\omega(b,c).$$

From this relation, with $a = x(k)$, $b = x(k+1)$, and $c = z$, we have

$$\begin{aligned}\Big\langle \nabla\omega\big(x(k)\big) - \nabla\omega\big(x(k+1)\big), z - x(k+1)\Big\rangle& \\ = D_\omega\big(x(k), x(k+1)\big) - D_\omega\big(x(k), z\big) &+ D_\omega\big(x(k+1), z\big).\end{aligned}$$

Substituting the preceding equality into (6.12) and re-arranging terms result in

$$\begin{aligned}\Psi(x(k+1)) - \Psi(z) + \frac{1}{\gamma(k)}D_\omega\big(x(k+1), z\big) \leq \Big\langle g_{\text{ave}}\big(d(k)\big), z - x(k+1)\Big\rangle &+ \frac{1}{\gamma(k)}D_\omega\big(x(k), z\big)\\ &- \frac{1}{\gamma(k)}D_\omega\big(x(k), x(k+1)\big)\\ &- \frac{\mu_\Psi}{2}\big\|z - x(k+1)\big\|^2.\end{aligned}$$

Since the distance generating function $\omega(x)$ is 1-strongly convex, we have the lower bound

$$D_\omega\big(x(k), x(k+1)\big) \geq \frac{1}{2}\|x(k+1) - x(k)\|^2,$$

which implies that

$$\begin{aligned}\Psi(x(k+1)) - \Psi(z) + \frac{1}{\gamma(k)}D_\omega\big(x(k+1), z\big) \leq \Big\langle g_{\text{ave}}\big(d(k)\big), z - x(k+1)\Big\rangle &+ \frac{1}{\gamma(k)}D_\omega\big(x(k), z\big)\\ &- \frac{1}{2\gamma(k)}\|x(k+1) - x(k)\|^2\\ &- \frac{\mu_\Psi}{2}\big\|z - x(k+1)\big\|^2. \qquad (6.13)\end{aligned}$$

The essential idea in the rest of the proof is to use convexity and smoothness of the expectation function f to bound $f(x(k+1)) - f(z)$ for each $z \in \operatorname{dom}\Psi$ and each $k \in \mathbb{N}_0$. According to Assumption 6.2, $\nabla F(x,\xi)$ and, hence, $\nabla f(x)$ are Lipschitz continuous with the constant L. By using the L-Lipschitz continuity of ∇f and then the convexity of f, we have

$$\begin{aligned}f(x(k+1)) &\leq f(x(d(k))) + \langle \nabla f(x(d(k))), x(k+1) - x(d(k))\rangle + \frac{L}{2}\|x(k+1) - x(d(k))\|^2\\ &\leq f\big(z\big) + \big\langle \nabla f\big(x(d(k))\big), x(k+1) - z\big\rangle + \frac{L}{2}\big\|x(k+1) - x(d(k))\big\|^2, \qquad (6.14)\end{aligned}$$

for any $z \in \operatorname{dom}\Psi$. Combining inequalities (6.13) and (6.14), and recalling that $\varphi(x) = f(x) + \Psi(x)$, we obtain

$$\begin{aligned}\varphi(x(k+1)) - \varphi(z) + \frac{1}{\gamma(k)} D_\omega\big(x(k+1), z\big) \le{}& \big\langle \nabla f\big(x(d(k))\big) - g_{\text{ave}}\big(d(k)\big), x(k+1) - z\big\rangle \\ &+ \frac{1}{\gamma(k)} D_\omega\big(x(k), z\big) \\ &- \frac{1}{2\gamma(k)}\|x(k+1) - x(k)\|^2 - \frac{\mu_\Psi}{2}\big\|z - x(k+1)\big\|^2 \\ &+ \frac{L}{2}\|x(k+1) - x(d(k))\|^2.\end{aligned}$$

We now rewrite the above inequality in terms of the error

$$e\big(d(k)\big) = \nabla f\big(x(d(k))\big) - g_{\text{ave}}\big(d(k)\big),$$

as follows:

$$\begin{aligned}\varphi(x(k+1)) - \varphi(z) + \frac{1}{\gamma(k)} D_\omega\big(x(k+1), z\big) \le{}& \big\langle e\big(d(k)\big), x(k+1) - z\big\rangle + \frac{1}{\gamma(k)} D_\omega\big(x(k), z\big) \\ &- \frac{1}{2\gamma(k)}\|x(k+1) - x(k)\|^2 - \frac{\mu_\Psi}{2}\big\|z - x(k+1)\big\|^2 \\ &+ \frac{L}{2}\|x(k+1) - x(d(k))\|^2 \\ ={}& \underbrace{\big\langle e\big(d(k)\big), x(k+1) - x(k)\big\rangle}_{U_1} \\ &+ \big\langle e\big(d(k)\big), x(k) - z\big\rangle + \frac{1}{\gamma(k)} D_\omega\big(x(k), z\big) \\ &- \frac{1}{2\gamma(k)}\|x(k+1) - x(k)\|^2 - \frac{\mu_\Psi}{2}\big\|z - x(k+1)\big\|^2 \\ &+ \frac{L}{2}\underbrace{\|x(k+1) - x(d(k))\|^2}_{U_2}. \end{aligned} \tag{6.15}$$

We will seek upper bounds on the quantities U_1 and U_2. Let $\{\eta(k)\}_{k\in\mathbb{N}_0}$ be a sequence of positive numbers. For U_1, we have

$$\begin{aligned}U_1 &\le \left|\left\langle \frac{1}{\sqrt{\eta(k)}} e\big(d(k)\big), \sqrt{\eta(k)}\big(x(k+1) - x(k)\big)\right\rangle\right| \\ &\le \frac{1}{2\eta(k)}\big\|e\big(d(k)\big)\big\|_*^2 + \frac{\eta(k)}{2}\big\|x(k+1) - x(k)\big\|^2,\end{aligned} \tag{6.16}$$

where the second inequality follows from the Fenchel-Young inequality applied to the conjugate pair $\frac{1}{2}\|\cdot\|^2$ and $\frac{1}{2}\|\cdot\|_*^2$, *i.e.*,

$$\big|\langle a, b\rangle\big| \le \frac{1}{2}\big\|a\big\|_*^2 + \frac{1}{2}\big\|b\big\|^2.$$

We now turn to U_2. It follows from definition $\tau(k) = k - d(k)$ that

$$\begin{aligned} U_2 &= \big(k - d(k) + 1\big)^2 \left\| \sum_{j=0}^{k-d(k)} \frac{x(k-j) - x(k-j+1)}{k - d(k) + 1} \right\|^2 \\ &= \big(\tau(k) + 1\big)^2 \left\| \sum_{j=0}^{\tau(k)} \frac{x(k-j) - x(k-j+1)}{\tau(k) + 1} \right\|^2 . \end{aligned}$$

Then, by the convexity of the norm $\|\cdot\|$, we conclude that

$$\begin{aligned} U_2 &\leq \big(\tau(k) + 1\big) \sum_{j=0}^{\tau(k)} \big\| x(k-j) - x(k-j+1) \big\|^2 \\ &\leq \big(\tau_{\max} + 1\big) \sum_{j=0}^{\tau_{\max}} \big\| x(k-j) - x(k-j+1) \big\|^2, \end{aligned} \tag{6.17}$$

where the last inequality comes from our assumption that $\tau(k) \leq \tau_{\max}$ for all $k \in \mathbb{N}_0$. Substituting inequalities (6.16) and (6.17) into the bound (6.15) and simplifying yield

$$\begin{aligned} \varphi(x(k+1)) - \varphi(z) + \frac{1}{\gamma(k)} D_\omega\big(x(k+1), z\big) \leq{} & \frac{1}{2\eta(k)} \big\| e\big(d(k)\big) \big\|_*^2 \\ & + \big\langle e\big(d(k)\big), x(k) - z \big\rangle + \frac{1}{\gamma(k)} D_\omega\big(x(k), z\big) \\ & + \frac{L(\tau_{\max} + 1)}{2} \sum_{j=0}^{\tau_{\max}} \big\| x(k-j) - x(k-j+1) \big\|^2 \\ & - \frac{1}{2} \left(\frac{1}{\gamma(k)} - \eta(k) \right) \| x(k+1) - x(k) \|^2 \\ & - \frac{\mu_\Psi}{2} \big\| z - x(k+1) \big\|^2 . \end{aligned}$$

Setting $z = x^\star$, where $x^\star \in X^\star$, completes the proof. ∎

The next result follows from Lemma 6.1 by taking summation of the relations in (6.11).

Lemma 6.2. *Let Assumptions 6.1–6.5 hold. Assume also that* $\{\gamma(k)\}_{k \in \mathbb{N}_0}$ *is set to*

$$\gamma(k) = \frac{1}{\eta(k) + L(\tau_{\max} + 1)^2}, \quad k \in \mathbb{N}_0,$$

where $\eta(k)$ is positive for all k. Then, the iterates $\{x(k)\}_{k\in\mathbb{N}_0}$ produced by Algorithm 1 satisfy

$$\begin{aligned}\sum_{k=0}^{T-1}\left(\varphi(x(k+1))-\varphi^\star\right) \le \sum_{k=0}^{T-1}&\frac{1}{2\eta(k)}\left\|e\big(d(k)\big)\right\|_*^2\\
&+\sum_{k=0}^{T-1}\left\langle e\big(d(k)\big),x(k)-x^\star\right\rangle+\frac{1}{\gamma(0)}D_\omega\big(x(0),x^\star\big)\\
&+\sum_{k=0}^{T-1}\left(\frac{1}{\gamma(k+1)}-\frac{1}{\gamma(k)}\right)D_\omega\big(x(k+1),x^\star\big)\\
&-\frac{\mu_\Psi}{2}\sum_{k=0}^{T-1}\left\|x(k+1)-x^\star\right\|^2,\end{aligned}$$

for all $T\in\mathbb{N}$.

Proof. Applying Lemma 6.1 with

$$\eta(k)=\frac{1}{\gamma(k)}-L(\tau_{\max}+1)^2,$$

adding and subtracting $\gamma(k+1)^{-1}D_\omega\big(x(k+1),x^\star\big)$ to the left-hand side of (6.11), and re-arranging terms, we obtain

$$\begin{aligned}\varphi(x(k+1))-\varphi^\star+\frac{1}{\gamma(k+1)}D_\omega\big(x(k+1),x^\star\big)\le{}&\frac{1}{2\eta(k)}\left\|e\big(d(k)\big)\right\|_*^2\\
&+\left\langle e\big(d(k)\big),x(k)-x^\star\right\rangle+\frac{1}{\gamma(k)}D_\omega\big(x(k),x^\star\big)\\
&+\left(\frac{1}{\gamma(k+1)}-\frac{1}{\gamma(k)}\right)D_\omega\big(x(k+1),x^\star\big)\\
&+\frac{L(\tau_{\max}+1)}{2}\sum_{j=0}^{\tau_{\max}}\left\|x(k-j)-x(k-j+1)\right\|^2\\
&-\frac{L(\tau_{\max}+1)^2}{2}\|x(k+1)-x(k)\|^2\\
&-\frac{\mu_\Psi}{2}\left\|x(k+1)-x^\star\right\|^2.\end{aligned}$$

Summing the preceding inequality over $k=0,\dots,T-1$, $T\in\mathbb{N}$, yields

$$
\begin{aligned}
\sum_{k=0}^{T-1}\big(\varphi(x(k+1))-\varphi^\star\big)+\frac{1}{\gamma(T)}D_\omega\big(x(T),x^\star\big) &\le \sum_{k=0}^{T-1}\frac{1}{2\eta(k)}\big\|e\big(d(k)\big)\big\|_*^2 \\
&\quad+\sum_{k=0}^{T-1}\big\langle e\big(d(k)\big),x(k)-x^\star\big\rangle+\frac{1}{\gamma(0)}D_\omega\big(x(0),x^\star\big) \\
&\quad+\sum_{k=0}^{T-1}\left(\frac{1}{\gamma(k+1)}-\frac{1}{\gamma(k)}\right)D_\omega\big(x(k+1),x^\star\big) \\
&\quad+\frac{L(\tau_{\max}+1)}{2}\sum_{k=0}^{T-1}\sum_{j=0}^{\tau_{\max}}\big\|x(k-j)-x(k-j+1)\big\|^2 \\
&\quad-\frac{L(\tau_{\max}+1)^2}{2}\sum_{k=0}^{T-1}\|x(k+1)-x(k)\|^2 \\
&\quad-\frac{\mu_\Psi}{2}\sum_{k=0}^{T-1}\big\|x(k+1)-x^\star\big\|^2 \\
&\le \sum_{k=0}^{T-1}\frac{1}{2\eta(k)}\big\|e\big(d(k)\big)\big\|_*^2 \\
&\quad+\sum_{k=0}^{T-1}\big\langle e\big(d(k)\big),x(k)-x^\star\big\rangle+\frac{1}{\gamma(0)}D_\omega\big(x(0),x^\star\big) \\
&\quad+\sum_{k=0}^{T-1}\left(\frac{1}{\gamma(k+1)}-\frac{1}{\gamma(k)}\right)D_\omega\big(x(k+1),x^\star\big) \\
&\quad-\frac{\mu_\Psi}{2}\sum_{k=0}^{T-1}\big\|x(k+1)-x^\star\big\|^2, \qquad (6.18)
\end{aligned}
$$

where the second inequality used the facts

$$
\begin{aligned}
\sum_{k=0}^{T-1}\sum_{j=0}^{\tau_{\max}}\big\|x(k-j)-x(k-j+1)\big\|^2 &= \sum_{j=0}^{\tau_{\max}}\sum_{k=-j}^{T-j-1}\big\|x(k)-x(k+1)\big\|^2 \\
&= \sum_{j=0}^{\tau_{\max}}\sum_{k=0}^{T-j-1}\big\|x(k)-x(k+1)\big\|^2 \\
&\le \sum_{j=0}^{\tau_{\max}}\sum_{k=0}^{T-1}\big\|x(k)-x(k+1)\big\|^2 \\
&\le (\tau_{\max}+1)\sum_{k=0}^{T-1}\big\|x(k)-x(k+1)\big\|^2,
\end{aligned}
$$

and $x(k)=x(0)$ for all $k\le 0$. Dropping the second term on the left-hand side of (6.18) concludes the proof. ■

Lemma 6.3. *Let $\|\cdot\|$ be a norm over $\mathbb{R}^n$ and let $\|\cdot\|_\star$ be its dual norm. Let ω be a 1-strongly convex function with respect to $\|\cdot\|$ over $\mathbb{R}^n$. If $y_1,\ldots,y_b\in\mathbb{R}^n$ are mean* *zero random variables drawn i.i.d. from a distribution $\mathcal{P}$, then*

$$\mathbb{E}\left[\left\|\frac{1}{b}\sum_{i=1}^{b} y_i\right\|_*^2\right] \le \frac{c}{b^2}\sum_{i=1}^{b}\mathbb{E}\left[\|y_i\|_*^2\right],$$

where $c \in [1, b]$ is given by

$$c = \begin{cases} 1, & \text{if } \|\cdot\|_* = \|\cdot\|_2, \\ 2\max_{\|x\|=1}\omega(x), & \text{otherwise.} \end{cases}$$

Proof. The result follows from [171, Lemma B.2] and convexity of the norm $\|\cdot\|_*$. For further details, see [150, §4.1]. ■

6.5.1 Proof of Theorem 6.1

Assume that the step-size $\{\gamma(k)\}_{k\in\mathbb{N}_0}$ is set to

$$\gamma(k) = \gamma = \frac{1}{\eta + L(\tau_{\max}+1)^2},$$

for some $\eta > 0$. It is clear that γ satisfies (6.8). Applying Lemma 6.2 with $\mu_\Psi = 0$, $\gamma(k) = \gamma$ and $\eta(k) = \eta$, we obtain

$$\sum_{k=0}^{T-1}\left(\varphi(x(k+1)) - \varphi^\star\right) \le \sum_{k=0}^{T-1}\frac{1}{2\eta}\left\|e\big(d(k)\big)\right\|_*^2 + \sum_{k=0}^{T-1}\left\langle e\big(d(k)\big), x(k) - x^\star\right\rangle + \frac{D_\omega\big(x(0), x^\star\big)}{\gamma}, \tag{6.19}$$

for all $T \in \mathbb{N}$. Each $x(k)$, $k \in \mathbb{N}$, is a deterministic function of the history $\xi_{[k-1]} := \{\xi_i(t) \mid i = 1, \ldots, b,\ t = 0, \ldots, k-1\}$ but not of $\xi_i(k)$. Since $\nabla f(x) = \mathbb{E}_\xi[\nabla_x F(x,\xi)]$, it follows that

$$\mathbb{E}_{|\xi_{[k-1]}}\left[\left\langle e\big(d(k)\big), x(k) - x^\star\right\rangle\right] = 0.$$

Moreover, as ξ_i and ξ_j are independent whenever $i \neq j$, it follows from Lemma 6.3 that

$$\begin{aligned}
\mathbb{E}\left[\|e\big(d(k)\big)\|_*^2\right] &= \mathbb{E}\left[\left\|\frac{1}{b}\sum_{i=1}^{b}\Big(\nabla f(x(d(k))) - \nabla_x F(x(d(k)),\xi_i)\Big)\right\|_*^2\right] \\
&\le \frac{c}{b^2}\sum_{i=1}^{b}\mathbb{E}\left[\|\nabla f(x(d(k))) - \nabla_x F(x(d(k)),\xi_i)\|_*^2\right] \\
&\le \frac{c\sigma^2}{b},
\end{aligned}$$

where the last inequality follows from Assumption 6.3. Taking expectation on both sides of (6.19) and using the above observations yield

$$\sum_{k=1}^{T}\big(\mathbb{E}[\varphi(x(k))]-\varphi^\star\big) \le \frac{c\sigma^2}{2\eta b}T + \frac{D_\omega\big(x(0),x^\star\big)}{\gamma}.$$

By the convexity of φ, we have

$$\varphi\big(x_{\text{ave}}(T)\big) = \varphi\left(\frac{1}{T}\sum_{k=1}^{T}x(k)\right) \le \frac{1}{T}\sum_{k=1}^{T}\varphi\big(x(k)\big),$$

which implies that

$$\mathbb{E}\big[\varphi(x_{\text{ave}}(T))\big]-\varphi^\star \le \frac{c\sigma^2}{2\eta b} + \frac{D_\omega\big(x(0),x^\star\big)}{\gamma T}.$$

Substituting $\eta = \gamma^{-1} - L(\tau_{\max}+1)^2$ into the above inequality proves the theorem.

6.5.2 Proof of Theorem 6.2

Assume that the step-size $\{\gamma(k)\}_{k\in\mathbb{N}_0}$ is chosen such that $\gamma(k)^{-1} = L(\tau_{\max}+1)^2+\alpha(k)$ where

$$\alpha(k) = \frac{\sigma\sqrt{c}\sqrt{k+1}}{R\sqrt{b}}.$$

Since $\gamma(k)$ is a non-increasing sequence, and $D_\omega(x,y) \le R^2$ for all $x,y \in \operatorname{dom}\Psi$, we have

$$\sum_{k=0}^{T-1}\left(\frac{1}{\gamma(k+1)}-\frac{1}{\gamma(k)}\right)D_\omega\big(x(k+1),x^\star\big) \le \left(\frac{1}{\gamma(T)}-\frac{1}{\gamma(0)}\right)R^2.$$

Applying Lemma 6.2 with $\mu_\Psi = 0$ and $\eta(k) = \alpha(k)$, taking expecation, and using Lemma 6.3 completely identically to the proof of Theorem 6.1, we then obtain

$$\sum_{k=1}^{T}\big(\mathbb{E}[\varphi(x(k))]-\varphi^\star\big) \le \frac{R^2}{\gamma(T)} + \frac{c\sigma^2}{2b}\sum_{k=0}^{T-1}\frac{1}{\alpha(k)}. \tag{6.20}$$

Viewing the sum as an lower-estimate of the integral of the function $y(t) = 1/\sqrt{t+1}$, one can verify that

$$\begin{aligned}\sum_{k=0}^{T-1}\frac{1}{\alpha(k)} = \sum_{k=0}^{T-1}\frac{1}{\widetilde{\alpha}\sqrt{k+1}} &\le \frac{1}{\widetilde{\alpha}}\left(1+\int_0^{T-1}\frac{dt}{\sqrt{t+1}}\right)\\ &\le \frac{2\sqrt{T}}{\widetilde{\alpha}},\end{aligned}$$

where $\widetilde{\alpha} = (\sigma\sqrt{c})/(R\sqrt{b})$. Substituting this inequality into the bound (6.20), we obtain the claimed guaranteed bound.

6.5.3 Proof of Theorem 6.3

Assume that $\{\gamma(k)\}_{k\in\mathbb{N}_0}$ is set to $\gamma(k)^{-1} = 2L(\tau_{\max}+1)^2 + \beta(k)$, with

$$\beta(k) = \frac{\mu_\Psi}{3Q}\big(k+\tau_{\max}+1\big).$$

We first describe some important properties of $\gamma(k)$ relevant to our proof. Clearly, $\gamma(k)$ is non-increasing, *i.e.*,

$$\frac{1}{\gamma(k)} \le \frac{1}{\gamma(k+1)}, \tag{6.21}$$

for all $k \in \mathbb{N}_0$. Since $\gamma(0)^{-1} \le \gamma(k)^{-1}$, we have

$$2L(\tau_{\max}+1)^2 + \frac{\mu_\Psi \tau_{\max}}{3Q} \le \frac{1}{\gamma(k)}. \tag{6.22}$$

Moreover, one can easily verify that

$$\begin{aligned}
\frac{1}{\gamma(k+1)^2} - \frac{1}{\gamma(k)^2} &= \frac{\mu_\Psi}{Q}\left(\frac{4L}{3}(\tau_{\max}+1)^2 + \frac{\mu_\Psi}{3Q}\left(\frac{2}{3}(k+\tau_{\max})+1\right)\right)\\
&\le \frac{\mu_\Psi}{Q}\left(2L(\tau_{\max}+1)^2 + \frac{\mu_\Psi}{3Q}\big(k+\tau_{\max}+1\big)\right)\\
&= \frac{\mu_\Psi}{Q}\frac{1}{\gamma(k)},
\end{aligned}$$

which implies that

$$\frac{1}{\gamma(k+1)^2} \le \frac{1}{\gamma(k)}\left(\frac{1}{\gamma(k)} + \frac{\mu_\Psi}{Q}\right), \tag{6.23}$$

for all $k \in \mathbb{N}_0$. Finally, by the definition of $\gamma(k)$, we have

$$\begin{aligned}
\frac{\gamma(k)}{\gamma(k+\tau_{\max})} &= 1 + \frac{\frac{\mu_\Psi}{3Q}\tau_{\max}}{2L(\tau_{\max}+1)^2 + \frac{\mu_\Psi}{3Q}\big(k+\tau_{\max}+1\big)}\\
&\le 1 + \frac{\mu_\Psi \tau_{\max}}{6LQ(\tau_{\max}+1)^2},
\end{aligned}$$

and hence,

$$\frac{1}{\gamma(k+\tau_{\max})} \le \left(1 + \frac{\mu_\Psi \tau_{\max}}{6LQ(\tau_{\max}+1)^2}\right)\frac{1}{\gamma(k)}. \tag{6.24}$$

We are now ready to prove Theorem 6.3. Applying Lemma 6.1 with

$$\eta(k) = \frac{1}{2\gamma(k)}, \quad k \in \mathbb{N}_0,$$

and using the fact

$$D_\omega\big(x(k+1), x^\star\big) \le \frac{Q}{2}\big\|x(k+1) - x^\star\big\|^2,$$

by Assumption 6.6, we obtain

$$\begin{aligned}
\varphi\big(x(k+1)\big) - \varphi^\star + \left(\frac{1}{\gamma(k)} + \frac{\mu_\Psi}{Q}\right) D_\omega\big(x(k+1), x^\star\big) \le \gamma(k)\big\|e\big(d(k)\big)\big\|_*^2 \\
+ \big\langle e\big(d(k)\big), x(k) - x^\star\big\rangle + \frac{1}{\gamma(k)} D_\omega\big(x(k), x^\star\big) \\
+ \frac{L(\tau_{\max}+1)}{2}\sum_{j=0}^{\tau_{\max}}\big\|x(k-j) - x(k-j+1)\big\|^2 \\
- \frac{1}{4\gamma(k)}\|x(k+1) - x(k)\|^2.
\end{aligned}$$

Multiplying both sides of this relation by $1/\gamma(k)$, and then using (6.23), we have

$$\begin{aligned}
\frac{1}{\gamma(k)}\big(\varphi\big(x(k+1)\big) - \varphi^\star\big) + \frac{1}{\gamma(k+1)^2} D_\omega\big(x(k+1), x^\star\big) \le \big\|e\big(d(k)\big)\big\|_*^2 \\
+ \frac{1}{\gamma(k)}\big\langle e\big(d(k)\big), x(k) - x^\star\big\rangle + \frac{1}{\gamma(k)^2} D_\omega\big(x(k), x^\star\big) \\
+ \frac{L(\tau_{\max}+1)}{2\gamma(k)}\sum_{j=0}^{\tau_{\max}}\big\|x(k-j) - x(k-j+1)\big\|^2 \\
- \frac{1}{4\gamma(k)^2}\|x(k+1) - x(k)\|^2.
\end{aligned}$$

Summing the above inequality from $k = 0$ to $k = T-1$, $T \in \mathbb{N}$, and dropping the first term on the left-hand side yield

$$\begin{aligned}
\frac{1}{\gamma(T)^2} D_\omega\big(x(T), x^\star\big) \le \sum_{k=0}^{T-1}\big\|e\big(d(k)\big)\big\|_*^2 \\
+ \sum_{k=0}^{T-1}\frac{1}{\gamma(k)}\big\langle e\big(d(k)\big), x(k) - x^\star\big\rangle + \frac{1}{\gamma(0)^2} D_\omega\big(x(0), x^\star\big) \\
+ \frac{L(\tau_{\max}+1)}{2}\sum_{k=0}^{T-1}\sum_{j=0}^{\tau_{\max}}\frac{1}{\gamma(k)}\big\|x(k-j) - x(k-j+1)\big\|^2 \\
- \frac{1}{4}\sum_{k=0}^{T-1}\frac{1}{\gamma(k)^2}\|x(k+1) - x(k)\|^2. \qquad (6.25)
\end{aligned}$$

What remains is to bound the third term on the right-hand side of (6.25). It follows from (6.21)–(6.24) that

$$
\begin{aligned}
&\frac{L(\tau_{\max}+1)}{2}\sum_{k=0}^{T-1}\sum_{j=0}^{\tau_{\max}}\frac{1}{\gamma(k)}\big\|x(k-j)-x(k-j+1)\big\|^2\\
&\qquad\qquad\qquad = \frac{L(\tau_{\max}+1)}{2}\sum_{j=0}^{\tau_{\max}}\sum_{k=0}^{T-j-1}\frac{1}{\gamma(k+j)}\big\|x(k)-x(k+1)\big\|^2\\
&\qquad\qquad\qquad \le \frac{L(\tau_{\max}+1)}{2}\sum_{j=0}^{\tau_{\max}}\sum_{k=0}^{T-1}\frac{1}{\gamma(k+j)}\big\|x(k)-x(k+1)\big\|^2\\
&\qquad\qquad\qquad \overset{(6.21)}{\le} \frac{L(\tau_{\max}+1)}{2}\sum_{j=0}^{\tau_{\max}}\sum_{k=0}^{T-1}\frac{1}{\gamma(k+\tau_{\max})}\big\|x(k)-x(k+1)\big\|^2\\
&\qquad\qquad\qquad = \frac{L(\tau_{\max}+1)^2}{2}\sum_{k=0}^{T-1}\frac{1}{\gamma(k+\tau_{\max})}\big\|x(k)-x(k+1)\big\|^2\\
&\qquad\qquad\qquad \overset{(6.24)}{\le} \frac{2L(\tau_{\max}+1)^2+\frac{\mu_\Psi\tau_{\max}}{3Q}}{4}\sum_{k=0}^{T-1}\frac{1}{\gamma(k)}\big\|x(k)-x(k+1)\big\|^2\\
&\qquad\qquad\qquad \overset{(6.22)}{\le} \frac{1}{4}\sum_{k=0}^{T-1}\frac{1}{\gamma(k)^2}\big\|x(k)-x(k+1)\big\|^2.
\end{aligned}
$$

Substituting the above inequality into (6.25), and then taking expectation on both sides (similarly to the proof of Theorems 6.1 and 6.2), we have

$$
\frac{1}{\gamma(T)^2}\mathbb{E}\big[D_\omega\big(x(T),x^\star\big)\big] \le \frac{c\sigma^2 T}{b} + \frac{1}{\gamma(0)^2}D_\omega\big(x(0),x^\star\big). \tag{6.26}
$$

As the distance generating function $\omega(x)$ is 1-strongly convex,

$$
\frac{1}{2}\|x(T)-x^\star\|^2 \le D_\omega\big(x(T),x^\star\big).
$$

Moreover, by the definition of $\gamma(k)$,

$$
\frac{\mu_\Psi(T+1)}{3Q} \le \beta(T) \le \frac{1}{\gamma(T)}.
$$

Combing these inequalities with the bound (6.26), we conclude

$$
\mathbb{E}\big[\|x(T)-x^\star\|^2\big] \le \frac{18c\sigma^2Q^2}{b\mu_\Psi^2(T+1)} + \frac{2\left(\frac{6LQ}{\mu_\Psi}+1\right)^2(\tau_{\max}+1)^4}{(T+1)^2}D_\omega\big(x(0),x^\star\big).
$$

The proof is complete.

Chapter 7

Conclusions and Future Work

In this chapter, we conclude the thesis by summarizing the main results and presenting some possible directions for future research.

7.1 Conclusions

In this thesis, we addressed several topics concerning the convergence analysis of positive nonlinear systems, contractive fixed-point iterations and stochastic optimization algorithms in the presence of time-delays. The main contributions are the following:

Delay-independent stability of positive systems: In Chapter 3, we extended a fundamental property of positive linear systems to a class of positive nonlinear systems. Specifically, we demonstrated that the stability of homogeneous positive monotone systems is independent of the magnitude and variation of time-varying delays. Since quantitative stability measures can be highly dependent on how fast the delays can grow large, we derived explicit expressions that allow us to quantify the impact of delays on the decay rate of homogeneous positive systems. We also showed that the best decay rate of positive linear systems that our results guarantee can be found by solving a tractable convex optimization problem.

Delay-independent stability of general positive monotone (not necessarily homogeneous) systems was discussed in Chapter 4. We presented a set of necessary and sufficient conditions for asymptotic stability of positive monotone systems with heterogeneous time-varying delays. We then proved that if a positive monotone system whose vector field is sub-homogeneous is globally asymptotically stable, then the corresponding system with bounded time-delays is also globally asymptotically stable. Furthermore, we used our results to analyze delay-independent stability of continuous-time power control algorithms in wireless networks.

Convergence rates of asynchronous contractive iterations: In Chapter 5, we studied the convergence of asynchronous fixed-point iterations involving maximum norm pseudo-contractions. We presented a powerful approach for characterizing the convergence rate of totally asynchronous iterations, where both the update intervals and communication delays may grow unbounded. Our approach also allows to explicitly quantify how the degree of asynchronism affects the convergence rate of partially asynchronous algorithms, where the update intervals and communication delays have a fixed upper bound. We demonstrated how our results can be used to analyze the impact of asynchrony on the convergence rate of discrete-time power control algorithms in wireless networks.

Asynchronous algorithm for stochastic optimization: Mini-batch optimization is a powerful paradigm for large-scale learning. However, the state-of-the-art parallel mini-batch algorithms assume synchronous operation or cyclic update orders. When worker nodes are heterogeneous (due to different computational capabilities or different communication delays), synchronous and cyclic operations are inefficient since they will leave workers idle waiting for the slower nodes to complete their computations. In Chapter 6, we proposed an asynchronous mini-batch algorithm for regularized stochastic optimization problems that eliminates idle waiting and allows workers to run at their maximal update rates. We showed that the algorithm achieves the rate $\mathcal{O}(1/\sqrt{T})$ for general convex regularization functions, and the rate $\mathcal{O}(1/T)$ for strongly convex regularization functions by suitably choosing the step-size values. In both cases, the impact of asynchrony on the convergence rate of our algorithm is asymptotically negligible, and a near-linear speedup in the number of workers can be expected.

7.2 Future work

There are several directions to further develop the work presented in the thesis. Some of them are discussed below.

Non-monotone positive systems: The results of Chapters 3 and 4 hold for positive nonlinear systems which are monotone. This may give the impression that the delay-independence property stems from monotonicity of such systems. Nonetheless, as shown in [128,172], the stability of particular classes of non-monotone positive systems are insensitive to time-delays. Extensions of our results to more general classes of positive nonlinear systems, for which the monotonicity assumption does not hold, is an interesting future topic to investigate.

Control design: In Chapter 3, we developed two theorems for global μ-stability of homogeneous positive systems that quantify the convergence rates for various classes of time-delays. It would be interesting to design controllers for time-delay homogeneous positive systems that guarantee that the resulting closed-loop system is positive and globally μ-stable with a desired decay rate.

Broad classes of asynchronous iterations: In Chapter 5, we considered fixed point iterations under contractivity assumption with respect to the maximum norm. It may be possible to derive convergence rate results for asynchronous iterations involving contraction mappings with respect to other norms, non-expansive mappings, and monotone mappings.

Non-i.i.d sampling: In order to establish our results in Chapter 6, we assumed that the stochastic oracle can generate i.i.d. samples from the distribution over which we optimize. This assumption is commonly used in the analysis for stochastic optimization algorithms, for example, [18–22, 154–156]. Recently, stochastic gradient methods for nonsmooth stochastic optimization were developed for situations in which there is no access to i.i.d. samples from the desired distribution [173]. Under reasonable assumptions on the ergodicity of the stochastic process that generates the samples, [173] obtained the convergence rate for serial mirror descent methods. It would be very interesting to extend this result to regularized stochastic optimization with smooth objective functions and investigate the convergence of asynchronous mini-batch algorithms when the random samples are dependent.

Accelerated asynchronous methods: For general convex regularization functions, the convergence rate of our asynchronous algorithm presented in Chapter 6 is

$$\mathcal{O}\left(\frac{L(\tau_{\max}+1)^2}{T}+\frac{\sigma}{\sqrt{T}}\right).$$

The first term is related to the smooth component in the objective function and the existence of time-delays in gradient computations while the second term is related to the variance in stochastic gradients. As mentioned in Section 6.1, the accelerated stochastic approximation method proposed in [19] can reduce the impact of the smooth component significantly in the absence of asynchrony and achieve the rate

$$\mathcal{O}\left(\frac{L}{T^2}+\frac{\sigma}{\sqrt{T}}\right).$$

Hence, an interesting question is whether an asynchronous version of this method decreases or increases the effects of time-delays on the convergence rate. Answering this question is, however, challenging and nontrivial, since the convergence analysis of accelerated first-order methods even in a deterministic and serial setting is much more involved than that of non-accelerated methods [85].

Bibliography

[1] L. Farina and S. Rinaldi. *Positive Linear Systems: Theory and Applications*. John Wiley and Sons, New York, 2000.

[2] E. Chauvet, J. E. Paullet, J. P. Previte, and Z. Walls. A Lotka-Volterra three-species food chain. *Mathematics Magazine*, pages 243–255, 2002.

[3] S. A. Grandhi, R. Vijayan, D. Goodman, and J. Zander. Centralized power control in cellular radio systems. *IEEE Transactions on Vehicular Technology*, 42(4):466–468, 1993.

[4] D. Angeli, P. D. Leenheer, and E. D. Sontag. On the structural monotonicity of chemical reaction networks. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 7–12, 2006.

[5] A. Berman and R. J. Plemmons. *Nonnegative Matrices in Mathematical Sciences*. Academic Press, New York, 1979.

[6] M. E. Valcher. On the internal stability and asymptotic behavior of 2-D positive systems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 44(7):602–613, 1997.

[7] P. D. Leenheer and D. Aeyels. Stabilization of positive linear systems. *Systems & Control Letters*, 44:259–271, 2001.

[8] M. A. Rami and F. Tadeo. Controller synthesis for positive linear systems with bounded controls. *IEEE Transactions on Circuits and Systems II*, 54(2):151–155, 2007.

[9] O. Mason and R. Shorten. On linear copositive Lyapunov functions and the stability of switched positive linear systems. *IEEE Transactions on Automatic Control*, 52(7):1346–1349, 2007.

[10] E. Fornasini and M. E. Valcher. Reachability of a class of discrete-time positive switched systems. *SIAM Journal on Control and Optimization*, 49(1):162–184, 2011.

[11] V. Jonsson, A. Rantzer, and R. M. Murray. A scalable formulation for engineering combination therapies for evolutionary dynamics of disease. *In Proceeding of IEEE American Control Conference (ACC)*, pages 2771–2778, 2014.

[12] M. Valcher and P. Misra. On the stabilizability and consensus of positive homogeneous multi-agent dynamical systems. *IEEE Transactions on Automatic Control*, 59(7):1936–1941, 2014.

[13] M. Colombino and R. S. Smith. A convex characterization of robust stability for positive and positively dominated linear systems. *IEEE Transactions on Automatic Control*, 2015.

[14] G. J. Foschini and Z. Miljanic. A simple distributed autonomous power control algorithm and its convergence. *IEEE Transactions on Vehicular Technology*, 42:641–646, 1993.

[15] R. Sipahi, S.-I. Niculescu, C. T. Abdallah, W. Michiels, and K. Gu. Stability and stabilization of systems with time delay. *IEEE Control Systems*, 31(1):38–65, 2011.

[16] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation.* Prentice-Hall, 1989.

[17] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, pages 400–407, 1951.

[18] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[19] G. Lan. An optimal method for stochastic composite optimization. *Mathematical Programming*, 133:365–397, 2012.

[20] C. Hu, W. Pan, and J. T. Kwok. Accelerated gradient methods for stochastic optimization and online learning. *Advances in Neural Information Processing Systems*, pages 781–789, 2009.

[21] S. Ghadimi and G. Lan. Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization I: A generic algorithmic framework. *SIAM Journal on Optimization*, 22(4):1469–1492, 2012.

[22] L. Xiao. Dual averaging method for regularized stochastic learning and online optimization. *Advances in Neural Information Processing Systems*, pages 2116–2124, 2009.

[23] M. Li, D. G. Andersen, A. J. Smola, and K. Yu. Communication efficient distributed machine learning with the parameter server. *Advances in Neural Information Processing Systems*, pages 19–27, 2014.

[24] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986.

[25] B. Recht and C. Ré. Parallel stochastic gradient algorithms for large-scale matrix completion. *Mathematical Programming Computation*, 5(2):201–226, 2013.

[26] M. Li, L. Zhou, Z. Yang, A. Li, F. Xia, D. G. Andersen, and A. J. Smola. Parameter server for distributed machine learning. *Big Learning NIPS Workshop*, 2013.

[27] M. Jaggi, V. Smith, M. Takác, J. Terhorst, S. Krishnan, T. Hofmann, and M. I. Jordan. Communication-efficient distributed dual coordinate ascent. *Advances in Neural Information Processing Systems*, pages 3068–3076, 2014.

[28] S. Chaturapruek, J. C. Duchi, and C. Ré. Asynchronous stochastic approximation: the noise is in the noise and SGD don't care. *Advances in Neural Information Processing Systems*, 2015.

[29] O. Fercoq and P. Richtárik. Accelerated, parallel and proximal coordinate descent. *SIAM Journal on Optimization*, 25(4):1997–2023, 2015.

[30] X. Lian, J. Liu, Y. Huang, and Y. Li. Asynchronous parallel stochastic gradient for nonconvex optimization. *Advances in Neural Information Processing Systems*, 2015.

[31] P. Richtárik and M. Takác. Parallel coordinate descent methods for big data optimization. *Mathematical Programming*, pages 1–52, 2015.

[32] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, and Q. V. Le. Large scale distributed deep networks. *Advances in Neural Information Processing Systems*, pages 1223–1231, 2012.

[33] T. Chilimbi, Y. Suzue, J. Apacible, and K. Kalyanaraman. Project Adam: Building an efficient and scalable deep learning training system. *11th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, pages 571–582, 2014.

[34] R. Zhang and J. Kwok. Asynchronous distributed ADMM for consensus optimization. *In Proceeding of International Conference on Machine Learning (ICML)*, pages 1701–1709, 2014.

[35] K. P. Bennett and O. L. Mangasarian. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1(1):23–34, 1992.

[36] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, pages 267–288, 1996.

[37] T. Hastie, R. Tibshirani, and J. Friedman. *Elements of Statistical Learning.* Springer, 2009.

[38] H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Asymptotic stability and decay rates of homogeneous positive systems with bounded and unbounded delays. *SIAM Journal on Control and Optimization*, 52(4):2623–2650, 2014.

[39] H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Exponential stability of homogeneous positive systems of degree one with time-varying delays. *IEEE Transactions on Automatic Control*, 59:1594–1599, 2014.

[40] H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Asymptotic stability and decay rates of positive linear systems with unbounded delays. *In Proceeding of IEEE Conference on Decision and Control Conference (CDC)*, pages 6337–6342, 2013.

[41] H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. On the rate of convergence of continuous-time linear positive systems with heterogeneous time-varying delays. *In Proceeding of European Control Conference (ECC)*, pages 3372–3377, 2013.

[42] H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Sub-homogeneous positive monotone systems are insensitive to heterogeneous time-varying delays. *In Proceeding of 21st International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, pages 317–324, 2014.

[43] H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Stability and performance of continuous-time power control in wireless networks. *IEEE Transactions on Automatic Control*, 59(8):2012–2023, 2014.

[44] H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Asymptotic and exponential stability of general classes of continuous-time power control laws in wireless networks. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 49–54, 2013.

[45] H. R. Feyzmahdavian and M. Johansson. On the convergence rates of asynchronous iterations. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 153–159, 2014.

[46] H. R. Feyzmahdavian, M. Johansson, and T. Charalambous. Contractive interference functions and rates of convergence of distributed power control laws. *IEEE Transactions on Wireless Communications*, 11(12):4494–4502, 2012.

[47] H. R. Feyzmahdavian, M. Johansson, and T. Charalambous. Contractive interference functions and rates of convergence of power control laws. *In Proceeding of IEEE International Conference on Communications (ICC)*, pages 5906–5910, 2012.

[48] H. R. Feyzmahdavian, A. Aytekin, and M. Johansson. An asynchronous mini-batch algorithm for regularized stochastic optimization. *Submitted to IEEE Transactions on Automatic Control*, 2015.

[49] H. R. Feyzmahdavian, A. Aytekin, and M. Johansson. An asynchronous mini-batch algorithm for regularized stochastic optimization. *To appear in IEEE Conference on Decision and Control (CDC)*, 2015.

[50] H. R. Feyzmahdavian, T. Charalambous, and M. Johansson. Delay-independent stability of cone-invariant monotone systems. *To appear in IEEE Conference on Decision and Control (CDC)*, 2015.

[51] B. Demirel, H. R. Feyzmahdavian, E. Ghadimi, and M. Johansson. Stability analysis of discrete-time linear systems with unbounded stochastic delays. *To appear in 5th IFAC Workshop on Distributed Estimation and Control of Networked Systems (NECSYS)*, 2015.

[52] E. Ghadimi, H. R. Feyzmahdavian, and M. Johansson. Global convergence of the heavy-ball method for convex optimization. *In Proceeding of European Control Conference (ECC)*, pages 310–315, 2015.

[53] J. Lu, H. R. Feyzmahdavian, and M. Johansson. A dual coordinate descent algorithm for multi-agent optimization. *In Proceeding of European Control Conference (ECC)*, pages 715–720, 2015.

[54] H. R. Feyzmahdavian, A. Aytekin, and M. Johansson. A delayed proximal gradient method with linear convergence rate. *In Proceeding of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2014.

[55] A. Aytekin, H. R. Feyzmahdavian, and M. Johansson. Asynchronous incremental block-coordinate descent. *In Proceeding of Annual Allerton Conference on Communication, Control, and Computing*, pages 19–24, 2014.

[56] H. R. Feyzmahdavian, A. Gattami, and M. Johansson. Distributed output-feedback LQG control with delayed information sharing. *In Proceeding of 3rd IFAC Workshop on Distributed Estimation and Control of Networked Systems (NECSYS)*, pages 192–197, 2012.

[57] H. R. Feyzmahdavian, A. Alam, and A. Gattami. Optimal distributed controller design with communication delays: Application to vehicle formations. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 2232–2237, 2012.

[58] W. M. Haddad, V. Chellaboina, and Q. Hui. *Nonnegative and Compartmental Dynamical Systems.* Princeton, 2010.

[59] F. Christiansen and T. M. Fenchel. *Theories of Populations in Biological Communities.* Springer Science & Business Media, 2012.

[60] H. K. Khalil. *Nonlinear Systems*, volume 3. Prentice Hall, NJ, 2002.

[61] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15. SIAM, 1994.

[62] A. Rantzer. Scalable control of positive systems. *European Journal of Control*, 2015.

[63] T. Tanaka and C. Langbort. The bounded real lemma for internally positive systems and H-Infinity structured static state feedback. *IEEE Transactions on Automatic Control*, 56(9):2218–2223, 2011.

[64] A. Rantzer. Distributed control of positive systems. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 6608–6611, 2011.

[65] Y. Ebihara, D. Peaucelle, and D. Arzelier. L_1 gain analysis of linear positive systems and its application. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 4029–4034, 2011.

[66] M. Colombino and R. S. Smith. Convex characterization of robust stability analysis and control synthesis for positive linear systems. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 4379–4384, 2014.

[67] C. Briat. Robust stability and stabilization of uncertain linear positive systems via integral linear constraints: L_1-gain and L_∞-gain characterization. *International Journal of Robust and Nonlinear Control*, 23(17):1932–1954, 2013.

[68] V. Blondel and J. N. Tsitsiklis. NP-hardness of sofme linear control design problems. *SIAM Journal on Control and Optimization*, 35(6):2118–2127, 1997.

[69] H. Smith. *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems.* American Mathematical Society, 1995.

[70] H. L. Smith. Systems of ordinary differential equations which generate an order preserving flow. A survey of results. *SIAM review*, 30(1):87–113, 1988.

[71] M. W. Hirsch. Systems of differential equations which are competitive or cooperative: I. limit sets. *SIAM Journal on Mathematical Analysis*, 13(2):167–179, 1982.

[72] M. W. Hirsch. Systems of differential equations that are competitive or cooperative: II. convergence almost everywhere. *SIAM Journal on Mathematical Analysis*, 16(3):423–439, 1985.

[73] M. W. Hirsch. Systems of differential equations which are competitive or cooperative: III. competing species. *Nonlinearity*, 1(1):51, 1988.

[74] D. Angeli and E. D. Sontag. Monotone control systems. *IEEE Transactions on Automatic Control*, 48(10):1684–1698, 2003.

[75] P. D. Leenheer, D. Angeli, and E. D. Sontag. Monotone chemical reaction networks. *Journal of mathematical chemistry*, 41(3):295–314, 2007.

[76] A. Aswani and C. Tomlin. Monotone piecewise affine systems. *IEEE Transactions on Automatic Control*, 54(8):1913–1918, 2009.

[77] A. Rantzer and B. Bernhardsson. Control of convex-monotone systems. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 2378–2383, 2014.

[78] E. Lovisari, G. Como, and K. Savla. Stability of monotone dynamical flow networks. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 2384–2389, 2014.

[79] D. Aeyels and P. D. Leenheer. Extension of the Perron-Frobenius theorem to homogeneous systems. *SIAM Journal on Control and Optimization*, 41(2):563–582, 2002.

[80] O. Mason and M. Verwoerd. Observations on the stability of cooperative systems. *Systems & Control Letters*, 58:461–467, 2009.

[81] P. D. Leenheer and D. Aeyels. Stability properties of equilibria of classes of cooperative systems. *IEEE Transactions on Automatic Control*, 46(12):1996–2001, 2001.

[82] D. Luenberger. *Introduction to Dynamic Systems: Theory, Models, and Applications.* Wiley, 1979.

[83] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar. *Convex Analysis and Optimization.* Athena Scientific, 2003.

[84] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge university press, 2004.

[85] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course.* Springer, 2004.

[86] L. M. Rios and N. V. Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293, 2013.

[87] A. S. Nemirovski and D. B. Yudin. *Problem complexity and method efficiency in optimization Problem.* John Wiley, 1983.

[88] A. Ben-Tal, T. Margalit, and A. Nemirovski. The ordered subsets mirror descent optimization method with applications to tomography. *SIAM Journal on Optimization*, 12(1):79–108, 2001.

[89] A. Beck and M. Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

[90] J. C. Duchi, S. Shalev-Shwartz, Y. Singer, and A. Tewari. Composite objective mirror descent. *In Annual Conference on Learning Theory (COLT)*, 2010.

[91] P. Tseng. Approximation accuracy, gradient methods, and error bound for structured convex optimization. *Mathematical Programming*, 125(2):263–295, 2010.

[92] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms.* Cambridge University Press, 2014.

[93] Y. Nesterov. A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. *Soviet Mathematics Doklady*, 27:372–376, 1983.

[94] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming*, 103(1):127–152, 2005.

[95] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[96] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. *submitted to SIAM Journal on Optimization.*

[97] Y. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013.

[98] R. D. Driver. *Ordinary and Delay Differential Equations.* Springer, New York, 1977.

[99] J. K. Hale and S. M. V. Lunel. *Introduction to Functional Differential Equations.* Springer, New York, 1993.

[100] E. Fridman and U. Shaked. An improved stabilization method for linear time-delay systems. *IEEE Transactions on Automatic Control*, 47(11):1931–1937, 2002.

[101] A. Seuret and F. Gouaisbaut. Wirtinger-based integral inequality: application to time-delay systems. *Automatica*, 49(9):2860–2866, 2013.

[102] E. Summers, M. Arcak, and A. Packard. Delay robustness of interconnected passive systems: an integral quadratic constraint approach. *IEEE Transactions on Automatic Control*, 58(3):712–724, 2013.

[103] E. Fridman. *Introduction to Time-Delay Systems: Analysis and Control*. Springer, 2014.

[104] W. M. Haddad and V. Chellaboina. Stability theory for non-negative and compartmental dynamical systems with time delay. *Systems & Control Letters*, 51(5):355–361, 2004.

[105] P. H. A. Ngoc. A Perron–Frobenius theorem for a class of positive quasi-polynomial matrices. *Applied Mathematics Letters*, 19(8):747–751, 2006.

[106] M. Buslowicz. Simple stability conditions for linear positive discrete time systems with delays. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 56:325–328, 2008.

[107] X. Liu, W. Yu, and L. Wang. Stability analysis of positive systems with bounded time-varying delays. *IEEE Transactions on Circuits and Systems II*, 56(7):600–604, 2009.

[108] M. A. Rami. Stability analysis and synthesis for linear positive systems with time-varying delays. *3rd International Symposium on Positive Systems (POSTA)*, pages 205–216, 2009.

[109] X. Liu, W. Yu, and L. Wang. Stability analysis for continuous-time positive systems with time-varying delays. *IEEE Transactions on Automatic Control*, 55(4):1024–1028, 2010.

[110] V. S. Bokharaie, O. Mason, and M. Verwoerd. D-stability and delay-independent stability of homogeneous cooperative systems. *IEEE Transactions on Automatic Control*, 55(12):2882–2885, 2010.

[111] X. Liu and C. Dang. Stability analysis of positive switched linear systems with delays. *IEEE Transactions on Automatic Control*, 56(7):1684–1690, 2011.

[112] T. Chen and L. Wang. Global μ-stability of delayed neural networks with unbounded time-varying delays. *IEEE Transactions on Neural Networks*, 18(6):1836–1840, 2007.

[113] A. Zappavigna, T. Charalambous, and F. Knorn. Unconditional stability of the Foschini-Miljanic algorithm. *Automatica*, 48(1):219–224, 2012.

[114] T. Chen and L. Wang. Power-rate global stability of dynamical systems with unbounded time-varying delays. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 54(8):705–709, 2007.

[115] B. Liu, W. Lu, and T. Chen. Generalized Halanay inequalities and their applications to neural networks with unbounded time-varying delays. *IEEE Transactions on Neural Networks*, 22(9):1508–1513, 2011.

[116] J. Shen and J. Lam. l_∞/L_∞-gain analysis for positive linear systems with unbounded time-varying delays. *IEEE Transactions on Automatic Control*, 60(3):857–862, 2015.

[117] H. Boche and M. Schubert. The structure of general interference functions and applications. *IEEE Transactions on Information Theory*, 54(11):4980–4990, 2008.

[118] D. G. Luenberger. *Introduction to Linear and Nonlinear Programming.* Addison-Wesley: Reading, MA, 1973.

[119] Y. Sun. Delay-independent stability of switched linear systems with unbounded time-varying delays. *Abstract and Applied Analysis*, 2012.

[120] X. Liu and J. Lam. Relationships between asymptotic stability and exponential stability of positive delay systems. *International Journal of General Systems*, 42(2):224–238, 2013.

[121] Y. Hino, S. Murakami, and T. Naito. *Functional Differential Equations with Infinite Delay.* Springer, New York, 1991.

[122] J. M. Danskin. The theory of Max–Min, with applications. *Journal on Applied Mathematics*, 14:641–664, 1966.

[123] B. S. Rüffer, C. M. Kellett, and S. R. Weller. Connection between cooperative positive systems and integral input-to-state stability of large-scale systems. *Automatica*, 46(6):1019–1027, 2010.

[124] S. Dashkovskiy, B. Rüffer, and F. Wirth. Discrete-time monotone systems: criteria for global asymptotic stability and applications. *17th Int. Symp. Math. Theory of Networks and Systems (MTNS)*, pages 89–97, 2006.

[125] J. D. Murray. *Mathematical Biology I: An Introduction.* Springer, New York, 2002.

[126] R. D. Yates. A framework for uplink power control in cellular radio systems. *IEEE Journal on Selected Areas in Communications*, 13(7):1341–1347, 1995.

[127] H. Boche and M. Schubert. A unifying approach to interference modeling for wireless networks. *IEEE Transactions on Signal Processing*, 58(6):3282–3297, 2010.

[128] V. S. Bokharaie and O. Mason. On delay-independent stability of a class of nonlinear positive time-delay systems. *IEEE Transactions on Automatic Control*, 59(7):1974–1977, 2014.

[129] V. S. Bokharaie, O. Mason, and F. Wirth. Stability and positivity of equilibria for subhomogeneous cooperative systems. *Nonlinear Analysis: Theory, Methods and Applications*, 74(17):6416–6426, 2011.

[130] W.-C. Su, B.-H. Jung, S.-Y. Chang, and Z. Gajic. Unified continuous- and discrete-time uplink power control problem formulation for SIR-based wireless networks. *IEEE Sarnoff Symposium*, pages 1–5, April 2009.

[131] T. Charalambous, I. Lestas, and G. Vinnicombe. On the stability of the Foschini-Miljanic algorithm with time-delays. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 2991–2996, 2008.

[132] C. A. Desoer and Y. Yang. On the generalized Nyquist stability criterion. *IEEE Transaction on Automatic Control*, 25(1):187–196, 1980.

[133] P. Tseng, D. P. Bertsekas, and J. N. Tsitsiklis. Partially asynchronous, parallel algorithms for network flow and other problems. *SIAM Journal on Control and Optimization*, 28(3):678–710, 1990.

[134] A. Frommer and D. B. Szyld. On asynchronous iterations. *Journal of Computational and Applied Mathematics*, 123(1):201–216, 2000.

[135] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 2996–3000, 2005.

[136] A. Nedić and A. Ozdaglar. Convergence rate for consensus with delays. *Journal of Global Optimization*, 47(3):437–456, 2010.

[137] D. P. Bertsekas and Huizhen Y. Distributed asynchronous policy iteration in dynamic programming. *In Proceeding of Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1368–1375, 2010.

[138] D. Chazan and W. Miranker. Chaotic relaxation. *Linear Algebra and its Applications*, 2(2):199–222, 1969.

[139] J. Miellou. Algorithmes de relaxation chaotique à retards. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 9(R1):55–82, 1975.

[140] G. M. Baudet. Asynchronous iterative methods for multiprocessors. *Journal of the ACM (JACM)*, 25(2):226–244, 1978.

[141] M. N. Tarazi. Some convergence results for asynchronous algorithms. *Numerische Mathematik*, 39(3):325–340, 1982.

[142] D. P. Bertsekas. Distributed dynamic programming. *IEEE Transactions on Automatic Control*, 27(3):610–616, 1982.

[143] D. P. Bertsekas and D. El Baz. Distributed asynchronous relaxation methods for convex network flow problems. *SIAM Journal on Control and Optimization*, 25(1):74–85, 1987.

[144] D. P. Bertsekas and J. Eckstein. Dual coordinate step methods for linear network flow problems. *Mathematical Programming*, 42(1):203–243, 1988.

[145] D. P. Bertsekas. Distributed asynchronous computation of fixed points. *Mathematical Programming*, 27(1):107–120, 1983.

[146] C. C. Moallemi and B. V. Roy. Convergence of min-sum message-passing for convex optimization. *IEEE Transactions on Information Theory*, 56:2041–2050, 2010.

[147] A. Yener, R. D. Yates, and S. Ulukus. CDMA multiuser detection: A nonlinear programming approach. *IEEE Transactions on Communications*, 50(6):1016–1024, 2002.

[148] M. Mehyar, D. Spanos, J. Pongsajapan, S. H. Low, and R. M. Murray. Asynchronous distributed averaging on communication networks. *IEEE/ACM Transactions on Networking*, 15:512–520, 2007.

[149] H. Avron, A. Druinsky, and A. Gupta. Revisiting asynchronous linear solvers: Provable convergence rate through randomization. *Parallel and Distributed Processing Symposium*, pages 198–207, 2014.

[150] O. Dekel, R. Gilad-Bachrach, O. Shamir, and L. Xiao. Optimal distributed online prediction using mini-batches. *Journal of Machine Learning Research*, 13(1):165–202, 2012.

[151] F. Niu, B. Recht, C. Ré, and S. J. Wright. Hogwild!: A lock-free approach to parallelizing stochastic gradient descent. *Advances in Neural Information Processing Systems*, pages 693–701, 2011.

[152] J. Liu, S. J. Wright, C. Ré, V. Bittorf, and S. Sridhar. An asynchronous parallel stochastic coordinate descent algorithm. *In Proceeding of International Conference on Machine Learning (ICML)*, pages 469–477, 2014.

[153] J. Liu and S. J. Wright. Asynchronous stochastic coordinate descent: Parallelism and convergence properties. *SIAM Journal on Optimization*, 25(1):351–376, 2015.

[154] A. Agarwal and J. C. Duchi. Distributed delayed stochastic optimization. *In Proceeding of IEEE Conference on Decision and Control (CDC)*, pages 5451–5452, 2012.

[155] S. Ghadimi and G. Lan. Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization, II: shrinking procedures and optimal algorithms. *SIAM Journal on Optimization*, 23(4):2061–2089, 2013.

[156] A. Rakhlin, O. Shamir, and K. Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. *In Proceeding of International Conference on Machine Learning (ICML)*, pages 449–456, 2012.

[157] A. Nedić and S. Lee. On stochastic subgradient mirror-descent algorithm with weighted averaging. *SIAM Journal on Optimization*, 24(1):84–107, 2014.

[158] R. T. Rockafellar and R. J. B. Wets. On the interchange of subdifferentiation and conditional expectation for convex functionals. *Stochastics: An International Journal of Probability and Stochastic Processes*, 7(3):173–182, 1982.

[159] Y. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical programming*, 120(1):221–259, 2009.

[160] X. Chen, Q. Lin, and J. Pena. Optimal regularized dual averaging methods for stochastic optimization. *Advances in Neural Information Processing Systems*, pages 395–403, 2012.

[161] M. Zinkevich, M. Weimer, L. Li, and A. J. Smola. Parallelized stochastic gradient descent. *Advances in Neural Information Processing Systems*, pages 2595–2603, 2010.

[162] J. C. Duchi, P. L. Bartlett, and M. J. Wainwright. Randomized smoothing for stochastic optimization. *SIAM Journal on Optimization*, 22(2):674–701, 2012.

[163] A. Nedić, D. P. Bertsekas, and V. S. Borkar. Distributed asynchronous incremental subgradient methods. *Studies in Computational Mathematics*, 8:381–407, 2001.

[164] J. C. Duchi, M. I. Jordan, and B. McMahan. Estimation, optimization, and parallelism when data is sparse. *Advances in Neural Information Processing Systems*, pages 2832–2840, 2013.

[165] B. McMahan and M. Streeter. Delay-tolerant algorithms for asynchronous distributed online learning. *Advances in Neural Information Processing Systems*, pages 2915–2923, 2014.

[166] J. Langford, A. J. Smola, and M. Zinkevich. Slow learners are fast. *Advances in Neural Information Processing Systems*, pages 2331–2339, 2009.

[167] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011.

[168] The Open MPI Project. Open MPI: Open Source High Performance Computing, 2015.

[169] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li. RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5:361–397, 2004.

[170] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

[171] A. Cotter, O. Shamir, N. Srebro, and K. Sridharan. Better mini-batch algorithms via accelerated gradient methods. *Advances in Neural Information Processing Systems*, pages 1647–1655, 2011.

[172] C. W. Sung and K. K. Leung. A generalized framework for distributed power control in wireless networks. *IEEE Transactions on Information Theory*, 51(7):2625–2635, 2005.

[173] J. C. Duchi, A. Agarwal, M. Johansson, and M. I. Jordan. Ergodic mirror descent. *SIAM Journal on Optimization*, 22(4):1549–1578, 2012.