Analysis and Synthesis of Switched Optimization Algorithms

Jared Miller[a], Fabian Jakob[b], Carsten Scherer[a], Andrea Iannelli[b]

[a] *Chair of Mathematical Systems Theory, Department of Mathematics, University of Stuttgart, Pfaffenwaldring 5a, Stuttgart, Germany70569,*
[b] *Institute for Systems Theory and Automatic Control, University of Stuttgart, Pfaffenwaldring 9, Stuttgart, Germany70569,*

Abstract

The deployment of optimization algorithms over communication networks face challenges associated with time delays and corruptions. Fixed time delays can destabilize popular gradient-based algorithms, and this degradation is exacerbated by time-varying delays that may arise from packet drops. This work concentrates on the analysis and synthesis of discrete-time optimization algorithms with certified exponential convergence rates that are robust against switched network dynamics between the optimizer and the gradient oracle. Analysis is accomplished by solving linear matrix inequalities under bisection in the exponential convergence rate and searching over Zames-Falb filter coefficients that can certify convergence. Synthesis is performed by alternating between a search over filter coefficient for a fixed controller, and a search over controllers for a fixed filter. Effectiveness is demonstrated by the synthesis of convergent optimization algorithms over networks with time-varying delays, and networks with unstable channel dynamics.

Keywords: Convex Optimization, Networks, Linear Matrix Inequalities, Switched Systems, Algorithms

Email addresses: jared.miller@imng.uni-stuttgart.de (Jared Miller), fabian.jakob@ist.uni-stuttgart.de (Fabian Jakob), carsten.scherer@imng.uni-stuttgart.de (Carsten Scherer), andrea.iannelli@ist.uni-stuttgart.de (Andrea Iannelli)

1. Introduction

Optimization algorithms increasingly operate within networked environments, where information exchange is rarely perfect. Communication links introduce delays, packet drops, and other distortions that alter how optimization algorithms interact with gradients. These effects can have significant consequences – slowing down convergence, degrading performance, or even destabilizing the algorithm dynamics.

In the strongly convex setting, robust control techniques have been used to analyze and synthesize optimization algorithms. Introduced by [1], the essential idea is to treat a first-order algorithm as the interconnection of a linear system with a static gradient nonlinearity, and to frame the convergence as an absolute stability problem of the resultant Lur'e system [2]. Algorithms should be designed such that they are applicable to a class of objective functions. Input-output relations satisfied by sequences arising from gradients in the function class are described using Integral Quadratic Constraints (IQCs) [3, 4], and the analysis and design of robust algorithms can subsequently be cast as a robust control problem [5]. This methodology also enables a quantification of robustness to noise in the gradients as measured by H_2 performance [6, 7], or to account for adversarial communication dynamics by considering them within an extremum control setup [8].

Most of this system-theoretic approach has focused on algorithm and communication dynamics that are each time-invariant [5, 9]. Even though recently this approach was extended to time-varying objectives and algorithms in [10], important network-induced phenomena such as *time-varying delays* have not been specifically addressed yet. When such effects are taken into account, the resulting dynamics are more accurately captured by switched systems [11, 12]. Motivated by this perspective, this work extends the system-theoretic framework to include *switched optimization algorithms*. The design of optimization algorithms in a switched setting has mostly focused on analysis and synthesis of distributed algorithms that are robust

against time-varying information delays in the form of stale (time-delayed) gradients [13, 14]. Delay-scheduled algorithms, in analogy to delay-scheduled controllers [15], have not yet been addressed.

This work performs analysis of switched optimization algorithms by solving mode-dependent [16] linear matrix inequalities (LMIs) [17]. The switched optimization synthesis problem is posed as a robustness requirement under a regulation constraint [18, 19], that is compatible under mode changes [20, 21]. Obedience of the regulation constraint is ensured through introduction of a network-dependent internal model. Given fixed Zames-Falb filters from the analysis procedure, we perform switched optimization algorithm synthesis by applying controller design techniques for LPV systems [22]. The convergence rates and empirical performance of the resulting algorithms are demonstrated on a network with unstable channel dynamics, and a network with time-varying delays.

The contributions of this work are as follows:

- A switched-system framework for first-order algorithms with bilateral corrupted gradient transmission, encompassing dynamical network models such as packet dropout and bounded-rate delay.
- An analysis procedure for upper-bounding worst-case convergence rates of given switched optimization algorithms.
- An internal-model-control-based synthesis procedure in the polytopic LPV setting for minimal worst-case linear convergence under switched network dynamics.

2. Preliminaries

We review preliminaries of notation and switched systems.

2.1. Notation

The set of natural numbers including 0 is $\mathbb{N}$. The set of natural numbers between a and b is $\{a, \dots, b\}$. An $\mathbb{N}$-indexed $\mathbb{R}^n$-valued signal is a map x :

$\mathbb{N} \to \mathbb{R}^n$ denoted as $\{x_k\}_{k\in\mathbb{N}}$. The 2-norm of a vector $g \in \mathbb{R}^n$ is $\|g\|_2 = \sqrt{g^\top g}$. A system G is a mapping that takes an initial condition $x_0 \in \mathbb{R}^n$ and an input signal $u : \mathbb{N} \to \mathbb{R}^{n_u}$, and returns an output signal $y : \mathbb{N} \to \mathbb{R}^{n_y}$. The system is linear if G is a linear map: for all $i \in \{1, 2\}$ and constants $\alpha_i \in \mathbb{R}$ and tuples (x_{0i}, u_i, y_i) with $G(x_{0i}, u_i) = y_i$, it holds that $G(\alpha_1 x_{01} + \alpha_2 x_{02}, \alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 y_1 + \alpha_2 y_2$. The linear system $G : (x_0, u) \to y$ may be represented by state space matrices $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ as

$$G : \qquad \left(\frac{x_{k+1}}{y_k}\right) = \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right) \left(\frac{x_k}{u_k}\right). \tag{1}$$

The symbol $e_i \in \mathbb{R}^n$ will denote the standard unit vector. The set of symmetric matrices of dimension n is $\mathbb{S}^n$. Positive (semi)-definitness of a symmetric matrix M will be denoted as $M \succ 0$ ($M \succeq 0$). The identity matrix of dimension n is I_n. The zeros and all-ones matrices of dimension $m \times n$ are $0_{m\times n}$ and $1_{m\times n}$, respectively.

2.2. Switched Systems

A switched system with N_s modes, n states, n_u inputs, and n_y outputs is described by a tuple $(\Sigma, \mathcal{G})$, where Σ is the set of the N_s subsystems, and $\mathcal{G}$ is a directed, unweighted graph with N_s-vertices. Each subsystem Σ_r for $r \in \{1, \ldots, N_s\}$ is assumed to be described by

$$\Sigma_r : \quad \left(\frac{x_{k+1}}{y_k}\right) = \left(\begin{array}{c|c} \mathcal{A}_r & \mathcal{B}_r \\ \hline \mathcal{C}_r & \mathcal{D}_r \end{array}\right) \left(\frac{x_k}{u_k}\right). \tag{2}$$

The notation Σ_r will refer to the specific representation in (2) at each $r \in 1, \ldots, s$. The vertices $\mathcal{V}$ of the switching graph $\mathcal{G}$ are the modes $r \in \{1, \ldots, N_s\}$. The edges $\mathcal{E} : (r, r') \in \mathcal{V} \times \mathcal{V}$ of $\mathcal{G}$ encode possible switching transitions. The set of all infinite-length paths $\{s_k\}_{k\in\mathbb{N}}$ in $\mathcal{G}$ is denoted as $\mathrm{Path}(\mathcal{G})$, in which $\mathrm{Path}(\mathcal{G})$ is a subset of all maps: $\mathbb{N} \to \{1, \ldots, N_s\}$ with $(s_k, s_{k+1}) \in \mathcal{E}$ $\forall k \in \mathbb{N}$. The switched system $(\Sigma, \mathcal{G})$ can therefore be interpreted as a map $(\Sigma, \mathcal{G}) : (x_0, s, u) \to y$.

A trajectory of the switched system $(\Sigma, \mathcal{G})$ is a tuple $(x_k, u_k, y_k, s_k)_{k\in\mathbb{N}}$ such that the relation in (2) at each $k \in \mathbb{N}$ with subsystem $r = s_k$. This tuple is uniquely specified by an initial condition $x_0 \in \mathbb{R}^n$, an input sequence $u : \mathbb{N} \to \mathbb{R}^{n_u}$, and a switching sequence $s \in \text{Path}(\mathcal{G})$.

The switched system $(\Sigma, \mathcal{G})$ is bounded (Lyapunov Stable) if there exists a $\gamma > 0$ such that when $u = 0$, for all initial points $x_0 \in \mathbb{R}^n$ and switching sequences $s \in \text{Path}(\mathcal{G})$, the relation $\|x_k\|_2 \leq \gamma\|x_0\|_2$ is obeyed. The switched system is asymptotically stable if, when $u = 0$, it holds that $\lim_{k\to\infty} x_k = 0$ for all initial points $x_0 \in \mathbb{R}^n$ and switching sequences $s \in \text{Path}(\mathcal{G})$. The switched system rejects constant disturbances if it is asymptotically stable and $\lim_{k\to\infty} y_k = 0$ holds for any constant input u [18].

Given two linear systems $G_1 : (x_1, (q, z)) \to (p, w)$ and $G_2 : (x_2, (w, u)) \to (z, y)$, we define their star product $G_1 \star G_2 : ((x_1, x_2), q, u) \to (p, y)$ as the well-posed feedback interconnection of G_1 and G_2 along the common channels w and z as computed from given state-space representations of G_1 and G_2 [23, Page 267-269]. Given two *switched* systems $(\Sigma_1, \mathcal{G}) : (x_1, s, (q, z)) \to (p, w)$ and $(\Sigma_2, \mathcal{G}) : (x_2, s, (w, u)) \to (z, y)$ sharing the same transition graph $\mathcal{G}$, we denote the star product $(\Sigma_1 \star \Sigma_2, \mathcal{G}) : ((x_1, x_2), s, (q, u)) \to (p, y)$ as their feedback interconnection comprised of taking the star product between the per-mode linear systems as $(\Sigma_1 \star \Sigma_2)_r = \Sigma_{1,r} \star \Sigma_{2,r}$ $\forall r \in 1, \ldots, s$. The common graph $\mathcal{G}$ may be omitted in the presentation of switched systems to simplify notation.

3. Switched Optimization Algorithms

Given scalars $0 < m < L < \infty$, we define the class $\mathcal{S}_{m,L}$ as the set of functions $f : \mathbb{R}^c \to \mathbb{R}$ such that $f - \frac{m}{2}\|\cdot\|_2^2$ and $\frac{L}{2}\|\cdot\|_2^2 - f$ are both convex [24, Definition 1.1]. In particular, $\mathcal{S}_{m,L}$ includes the set of twice-continuously-differentiable functions f that have a bounded Hessian as $mI \preceq \nabla^2 f(z) \preceq LI$ for all $z \in \mathbb{R}^c$. The optimization problem under consideration is

$$z^* \in \operatorname*{argmin}_{z\in\mathbb{R}^c} f(z), \tag{3}$$

in which $f \in \mathcal{S}_{m,L}$. Because $m > 0$, the optimum z^* solving (3) is unique. The set $\mathcal{S}^0_{m,L} \subset \mathcal{S}_{m,L}$ is the class of functions $f \in \mathcal{S}_{m,L}$ such that $f(0) = 0$ and $\nabla f(0) = 0$; if $f \in \mathcal{S}^0_{m,L}$, then $z^* = 0$ satisfies (3).

An optimization algorithm is a causal procedure that returns a sequence of iterates $\{z_k\}_{k\in\mathbb{N}}$ given an initial iterate $z_0 \in \mathbb{R}^c$ and a function $f \in \mathcal{S}_{m,L}$. The optimization algorithm is globally *convergent* if for every initial iterate $z_0 \in \mathbb{R}^c$, the returned sequence of iterates $\{z_k\}_{k\in\mathbb{N}}$ satisfies $\lim_{k\to\infty} z_k = z^*$ where z^* satisfies (3).

3.1. Algorithm Model

The algorithmic interconnection of the gradient of a function $f \in \mathcal{S}_{m,L}$ and switched system $(\Sigma, \mathcal{G})$ with $\mathcal{D}_r = 0$ at all modes is

$$\left(\begin{array}{c} x_{k+1} \\ \hline z_k \end{array}\right) = \left(\begin{array}{c|c} \mathcal{A}_{s_k} & \mathcal{B}_{s_k} \\ \hline \mathcal{C}_{s_k} & 0 \end{array}\right) \left(\begin{array}{c} x_k \\ \hline w_k \end{array}\right), \qquad w_k = \nabla f(z_k). \tag{4}$$

A trajectory of the algorithmic interconnection (4) is a tuple $(x_k, w_k, z_k, s_k)_{k\in\mathbb{N}}$ satisfying (4) for all $k \in \mathbb{N}$.

A *fixed orbit* of (4) is a tuple $(\{x^*_r\}_{r=1}^{N_s}, w^*, z^*)$ such that

$$\forall (r, r') \in \mathcal{E} : \left(\begin{array}{c} x^*_{r'} \\ \hline z^* \end{array}\right) = \left(\begin{array}{c|c} \mathcal{A}_r & \mathcal{B}_r \\ \hline \mathcal{C}_r & 0 \end{array}\right) \left(\begin{array}{c} x^*_r \\ \hline w^* \end{array}\right), \qquad w^* = \nabla f(z^*). \tag{5}$$

The interconnection in (4) is a *convergent switched optimization algorithm* if there exists a fixed orbit $(\{x^*_r\}_{r=1}^{N_s}, w^*, z^*)$ such that the following properties are satisfied:

1. **Optimality:** $w^* = 0$.
2. **Zero-State:** If $z^* = 0$, then $x^*_r = 0$ for all $r \in \{1, \ldots, N_s\}$.
3. **Attractivity:** Every trajectory (x_k, w_k, z_k, s_k) satisfies

$$\lim_{k\to\infty} \|x_k - x^*_{s_k}\|_2 = 0, \qquad \lim_{k\to\infty} \|z_k - z^*\|_2 = 0, \qquad \lim_{k\to\infty} w_k = 0. \tag{6}$$

3.2. Algorithms over Networks

The setting of switched networked optimization algorithms includes an algorithm (4) in which Σ is formed through the interconnection of a network $(P, \mathcal{G})$ with state x^N and a controller $(K, \mathcal{G})$ with state ξ. For each fixed mode $r \in \{1, \dots, N_s\}$, the algorithmic interconnection is described by

$$\nabla f : \qquad w_k = \nabla f(z_k), \tag{7a}$$

$$P_r : \qquad \left(\begin{array}{c} x_{k+1}^N \\ \hline z_k \\ y_k \end{array}\right) = \left(\begin{array}{c|cc} A_r & B_{r,1} & B_{r,2} \\ \hline C_{r,1} & D_{r,11} & D_{r,12} \\ C_{r,2} & D_{r,21} & D_{r,22} \end{array}\right) \left(\begin{array}{c} x_k^N \\ \hline w_k \\ u_k \end{array}\right), \tag{7b}$$

$$K_r : \qquad \left(\begin{array}{c} \xi_{k+1} \\ \hline u_k \end{array}\right) = \left(\begin{array}{c|c} A_r^K & B_r^K \\ \hline C_r^K & D_r^K \end{array}\right) \left(\begin{array}{c} \xi_k \\ \hline y_k \end{array}\right). \tag{7c}$$

The resulting system is the interconnection $(\Sigma, \mathcal{G}) = (P \star K, \mathcal{G})$.

An example of a switched optimization algorithm is a gradient descent scheme with stepsize γ under time varying delay h_k as described by the iteration $z_{k+1} = z_k - \gamma \nabla f(z_{k-h_k})$ for all $k \in \mathbb{N}$. The no-delay gradient descent algorithm has $h_k = 0$ for all $k \in \mathbb{N}$. Gradient descent under time-varying delay can therefore be understood as the interconnection of a time-varying-delay-describing network (P) and a no-delay gradient descent algorithm (K). These time-varying delays can arise in networked control systems for which an additional delay is incurred every time a packet transmission fails (packet drop). If the delay can take values between 0 and $h_{\max}$ inclusive, then the delay in a single coordinate can be mapped to a switched system [12] with modes $r \in \{1, \dots, h_{\max}+1\}$ as

$$\begin{aligned} A_r^d &= \begin{pmatrix} 0_{1 \times h_{\max}-1} & 0 \\ I_{h_{\max}-1} & 0_{h_{\max}-1 \times 1} \end{pmatrix}, \quad B_r^d = \begin{pmatrix} 1 \\ 0_{h_{\max}-1 \times 1} \end{pmatrix}, \\ C_r^d &= \begin{cases} 0_{1 \times h_{\max}-1} & r = 1 \\ e_{r-1}^\top & r \neq 1, \end{cases} \qquad D_r^d = \begin{cases} 1 & r = 1 \\ 0 & r \neq 1, \end{cases} \end{aligned} \tag{8}$$

in which the standard unit vector has dimensionality $e_{r-1} \in \mathbb{R}^{h_{\max}}$ for all $r \in \mathbb{N}$. A variable delay taking values in $h_k \in \{0, 1, 2\}$ over all coordinates can be represented as the switched system

$$\left(\begin{array}{c|c} A_r^d & B_r^d \\ \hline C_r^d & D_r^d \end{array}\right) \in \left\{ \left(\begin{array}{cc|c} 0 & 0 & I \\ I & 0 & 0 \\ \hline 0 & 0 & I \end{array}\right), \left(\begin{array}{cc|c} 0 & 0 & I \\ I & 0 & 0 \\ \hline I & 0 & 0 \end{array}\right), \left(\begin{array}{cc|c} 0 & 0 & I \\ I & 0 & 0 \\ \hline 0 & I & 0 \end{array}\right) \right\}. \tag{9}$$

If the delay can take the values of $h_k \in \{0, 1, 2\}$, the switched gradient descent scheme can be obtained using the delay primitives in (9) through the following interconnection for each $r \in \{1, 2, 3\}$:

$$\Sigma_r = P_r \star K_r, \qquad \left(\begin{array}{c|c} \mathcal{A}_r & \mathcal{B}_r \\ \hline \mathcal{C}_r & \mathcal{D}_r \end{array}\right) = \left(\begin{array}{c|c:c} A_r^d & 0 & B_r^d \\ \hline C_r^d & 0 & D_r^d \\ \hdashline 0 & I & 0 \end{array}\right) \star \left(\begin{array}{c|c} I & -\gamma I \\ \hline I & 0 \end{array}\right). \tag{10}$$

The subsystems Σ_r of the above interconnection admit the following description:

$$\left(\begin{array}{c|c} \mathcal{A}_r & \mathcal{B}_r \\ \hline \mathcal{C}_r & \mathcal{D}_r \end{array}\right) \in \left\{ \left(\begin{array}{ccc|c} 0 & 0 & I & I \\ I & 0 & 0 & 0 \\ 0 & 0 & I & -\gamma I \\ \hline 0 & 0 & I & 0 \end{array}\right), \left(\begin{array}{ccc|c} 0 & 0 & I & I \\ I & 0 & 0 & 0 \\ 0 & 0 & I & -\gamma I \\ \hline I & 0 & 0 & 0 \end{array}\right), \left(\begin{array}{ccc|c} 0 & 0 & I & I \\ I & 0 & 0 & 0 \\ 0 & 0 & I & -\gamma I \\ \hline 0 & I & 0 & 0 \end{array}\right) \right\} \tag{11}$$

Example graphs $\mathcal{G}$ associated with two time-varying delay models are shown in Figure 1. Figure 1a shows a graph that encodes delays taking values $h_k \in \{2, \ldots, 5\}$ with a bounded rate of change $|h_{k+1} - h_k| \leq 1$ for all $k \in \mathbb{N}$. The graph in Figure 1b describes the packet drop case as a reset system: the delay $h_k \in \{0, 1, 2, 3\}$ either increases by one every time step (packet drop) or drops to 0 (successful packet transmission). The maximum possible delay is 3 steps, upon which the next delay is 0.

4. Optimization Algorithm Analysis

We begin by providing a sufficient condition based on regulation theory under which switched optimization algorithms are convergent according to



(a) Bounded rate delay.

(b) Packet drop.

Figure 1: Examples of switching graphs $\mathcal{G}$ for different models of time-varying delay.

the definition in Section 3.1. We then use linear matrix inequality methods to upper-bound the exponential convergence rate of a given switched optimization algorithm.

4.1. Regulation of Switched Optimization Algorithms

We first describe a regulation property [18, 11] that is sufficient to ensure that trajectories of (4) converge to an optimal point z^*.

Theorem 4.1. *The interconnection (4) forms a convergent switched optimization algorithm if there exist matrices $(\Theta_r)_{r=1}^{N_s}$ such that*

$$\forall (r, r') \in \mathcal{E}: \qquad \mathcal{A}_r \Theta_r - \Theta_{r'} = 0, \tag{12a}$$

$$\mathcal{C}_r \Theta_r = I_c, \tag{12b}$$

and such that the following switched Luré interconnection is asymptotically stable for all $v_0 \in \mathbb{R}^n$, $f^0 \in \mathcal{S}_{m,L}^0$, $s \in Path(\mathcal{G})$:

$$v_{k+1} = \mathcal{A}_{s_k} v_k + \mathcal{B}_{s_k} \tilde{w}_k, \qquad \tilde{w}_k = \nabla f^0(\tilde{z}_k), \tag{13a}$$

$$\tilde{z}_k = \mathcal{C}_{s_k} v_k. \tag{13b}$$

Proof: Pick any $f \in \mathcal{S}_{m,L}$ and let z^* be the unique solution of the optimization problem (3). Then $w^* := \nabla f(z^*)$ satisfies $w^* = 0$, and hence also Property 1. in (6) is satisfied. If we define $x_r^* := \Theta_r z^*$ for $r \in \{1, \dots, N_s\}$, we can exploit (12) to infer

$$x_{r'}^* = \Theta_{r'} z^* = \mathcal{A}_r \Theta_r z^* = \mathcal{A}_r \Theta_r z^* + \mathcal{B}_r w^* \quad \text{and} \quad z^* = \mathcal{C}_r \Theta_r z^* = \mathcal{C}_r x_r^*$$

for all $(r, r') \in \mathcal{E}$. This shows that $(\{x_r^*\}_{r=1}^{N_s}, z^*, w^*)$ satisfies (5) and is a fixed orbit, for which Property 2. in (6) holds true.

Next, we pick any trajectory $(x_k, w_k, z_k, s_k)_{k\in\mathbb{N}}$ of (4). If we introduce the error signals

$$\tilde{x}_k := x_k - x^*_{s_k}, \qquad \tilde{z}_k := z_k - z^*, \qquad \tilde{w}_k := w_k - w^* = w_k, \tag{14}$$

for $k \in \mathbb{N}$, we need to prove $\lim_{k\to\infty}(\tilde{x}_k, \tilde{z}_k, \tilde{w}_k) = 0$ in order to ensure (6). We also define the shifted objective function $f^0 \in \mathcal{S}^0_{m,L}$ as

$$f^0(\zeta) := f(\zeta + z^*) \quad \text{for} \quad \zeta \in \mathbb{R}^c \tag{15}$$

to infer $\tilde{w}_k = w_k = \nabla f(z_k) = \nabla f(\tilde{z}_k + z^*) = \nabla f^0(\tilde{z}_k)$ for all $k \in \mathbb{N}$. Hence, the algorithmic interconnection (4) is easily rearranged to

$$\left(\begin{array}{c} x_{k+1} \\ \hdashline \tilde{z}_k \\ e_k \end{array}\right) = \left(\begin{array}{c|c:c} \mathcal{A}_{s_k} & 0 & \mathcal{B}_{s_k} \\ \hline \mathcal{C}_{s_k} & -I & 0 \\ \mathcal{C}_{s_k} & -I & 0 \end{array}\right) \left(\begin{array}{c} x_k \\ \hdashline z^* \\ \hdashline \tilde{w}_k \end{array}\right), \qquad \tilde{w}_k = \nabla f^0(\tilde{z}_k); \tag{16}$$

we introduce two copies of the output because the signal $\tilde{z}$ is the input to the nonlinearity ∇f^0 while e is interpreted as a regulated variable. Since the switching-dependent state error signal $\tilde{x}_k = x_k - x^*_{s_k}$ equals $x_k - \Theta_{s_k} z^*$ for all $k \in \mathbb{N}$, (16) transforms into

$$\left(\begin{array}{c} \tilde{x}_{k+1} \\ \hdashline \tilde{z}_k \\ e_k \end{array}\right) = \left(\begin{array}{c|c:c} \mathcal{A}_{s_k} & \mathcal{A}_{s_k}\Theta_{s_k} - \Theta_{s_{k+1}} & \mathcal{B}_{s_k} \\ \hline \mathcal{C}_{s_k} & \mathcal{C}_{s_k}\Theta_{s_k} - I & 0 \\ \mathcal{C}_{s_k} & \mathcal{C}_{s_k}\Theta_{s_k} - I & 0 \end{array}\right) \left(\begin{array}{c} \tilde{x}_k \\ \hdashline z^* \\ \hdashline \tilde{w}_k \end{array}\right), \quad \tilde{w}_k = \nabla f^0(\tilde{z}_k). \tag{17}$$

Now we note that $(s_k, s_{k+1}) \in \mathcal{E}$ for all $k \in \mathbb{N}$, which permits to exploit again (12) to conclude that (17) just reads as

$$\left(\begin{array}{c} \tilde{x}_{k+1} \\ \hdashline \tilde{z}_k \\ e_k \end{array}\right) = \left(\begin{array}{c|c:c} \mathcal{A}_{s_k} & 0 & \mathcal{B}_{s_k} \\ \hline \mathcal{C}_{s_k} & 0 & 0 \\ \mathcal{C}_{s_k} & 0 & 0 \end{array}\right) \left(\begin{array}{c} \tilde{x}_k \\ \hdashline z^* \\ \hdashline \tilde{w}_k \end{array}\right), \qquad \tilde{w}_k = \nabla f^0(\tilde{z}_k). \tag{18}$$

As a result, z^* is completely decoupled, and (18) is identical to (13). The assumed stability property ensures $\lim_{k\to\infty} \tilde{x}_k = 0$, and hence also $\lim_{k\to\infty} \tilde{z}_k = \lim_{k\to\infty} e_k = \lim_{k\to\infty} \mathcal{C}_{s_k}\tilde{x}_k = 0$ as well as $\lim_{k\to\infty} \tilde{w}_k = \lim_{k\to\infty} f^0(\tilde{z}_k) = 0$. Since $s \in \mathrm{Path}(\mathcal{G})$ was arbitrary, we have shown Property 3. in (6), which establishes that (4) is a convergent switched optimization algorithm. □

Remark 1. *Theorem 4.1 can be embedded into the regulation problem for switched linear systems in the literature by fixing the function $f \in \mathcal{S}_{m,L}$ in (4) to the very special case $f(z) = \frac{1}{2}m\|z - z^*\|^2$. The interconnection (4) with the error output $e_k := z_k - z^*$ then reads as*

$$x_{k+1} = (\mathcal{A}_{s_k} + m\mathcal{B}_{s_k}\mathcal{C}_{s_k})x_k - m\mathcal{B}_{s_k}z^* \quad \text{and} \quad e_k = \mathcal{C}_{s_k}x_k - z^*. \tag{19}$$

Regulation of linear systems amounts to ensuring that (19) is asymptotically stable for $z^* = 0$ and satisfies $\lim_{k\to\infty} e_k = 0$ for all $z \in \mathbb{R}^c$.

There exists a wide literature on guaranteeing uniform exponential stability of switched systems based on common quadratic or switched Lyapunov functions [25], which are amenable to the convex design of switching dependent controllers [22]. In particular, the work in [22] forms the basis for synthesis in the current paper.

For LTI systems, the additional regulation property is assured through the existence of a solution of the related regulator equation as known from the internal model principle [18, 19]. The extension to switched systems is handled in [26] and [27] by a common solution of the regulator equation; if applied for (19), this equation boils down to (12) with Θ being independent of the node index r. Motivated by almost regulation for LPV systems [20], our result is based on the use of mode-dependent solutions of the regulator equation to reduce conservatism and to even guarantee robust regulation against structured nonlinear uncertainties ∇f with $f \in \mathcal{S}_{m,L}$.

The use of path-dependent [28] or path-complete [29] Lyapunov functions as surveyed in [30] permits to arrive at necessary and sufficient conditions for uniform exponential stability (which are as well amenable to controller synthesis) by lifting the switched system into a larger system based on ℓ-length histories. To the best of our knowledge, an analogous result to ensure necessary and sufficient conditions for output regulation with path-dependent solutions of the regulator equation has not been considered in the literature.

4.2. Exponential Convergence Rate Analysis

Theorem 4.1 offers a condition for convergence. The speed of convergence can be characterized by ρ-exponential convergence. Formally, the interconnection (4) is a *ρ-exponentially convergent switched optimization algorithm* for some $\rho \in (0, 1)$, if it is a convergent switched optimization algorithm for

all $f \in \mathcal{S}_{m,L}$ and if there exists a $\gamma > 0$ such that all its trajectories satisfy

$$\|x_k - x^*_{s_k}\|_2 \leq \gamma \rho^k \|x_k - x^*_{s_0}\|_2 \qquad \forall k \in \mathbb{N}.$$

With $\mu := \max_{r\in\{1,\dots,N_s\}} \|\mathcal{C}_r\|$ (where $\|.\|$ is the spectral norm), we then observe that all trajectories of such an algorithm (4) also satisfy

$$\|z_k - z^*\|_2 = \|\mathcal{C}_{s_k} x_k - \mathcal{C}_{s_k} x^*_{s_k}\|_2 \leq \mu \|x_k - x^*_{s_k}\|_2 \leq \mu\gamma\rho^k \|x_0 - x^*_{s_0}\|_2 \tag{20}$$

for all $k \in \mathbb{N}$. Hence ρ-exponential algorithm convergence implies that z_k converges to z^* exponentially fast with rate ρ for $k \to \infty$.

The resulting convergence rate analysis problem is formulated as follows

Problem 1. *Given a function class $\mathcal{S}_{m,L}$, a switched system $(\Sigma, \mathcal{G})$, and a rate $\rho > 0$, certify whether the interconnection (4) is a ρ-convergent switched optimization algorithm.*

Theorem 4.1 is modified for ρ-exponential convergence as follows.

Corollary 1. *The interconnection (4) forms a ρ-exponentially convergent switched optimization algorithm if the condition (12) holds and the following system is bounded (Lyapunov stable):*

$$\bar{v}_{k+1} = (\rho^{-1}\mathcal{A}_{s_k})\bar{v}_k + (\rho^{-1}\mathcal{B}_{s_k})\bar{w}_k, \qquad \bar{w}_k = \rho^{-k}\nabla f^0(\rho^k \bar{z}_k), \tag{21a}$$

$$\bar{z}_k = \mathcal{C}_{s_k}\bar{v}_k. \tag{21b}$$

The proof follows arguments from [17] about exponential stability. Let $\mathcal{T}_{\rho^{-1}}$ denote ρ-exponential signal weighting, transforming a sequence v as

$$\mathcal{T}_{\rho^{-1}} : (v_0, v_1, v_2, \dots) \mapsto (v_0, \rho^{-1}v_1, \rho^{-2}v_2, \dots) =: (\bar{v}_0, \bar{v}_1, \bar{v}_2, \dots). \tag{22}$$

If we choose any trajectory $(v, \tilde{z}, \tilde{w})$ of (13), it is easily seen that the transformed signals $\bar{v} := \mathcal{T}_{\rho^{-1}}v$, $\bar{z} := \mathcal{T}_{\rho^{-1}}\tilde{z}$ and $\bar{w} := \mathcal{T}_{\rho^{-1}}\tilde{w}$ generate a trajectory of (21). Since the latter system is bounded, we infer $\|\bar{v}_k\|_2 \leq \gamma\|\bar{v}_0\|_2$ for all $k \in \mathbb{N}$ and some $\gamma > 0$, which translates into $\|v_k\|_2 \leq \gamma\rho^{-k}\|v_0\|_2$ for all $k \in \mathbb{N}$. This is the key modification to assure ρ-exponential algorithm convergence by following the proof of Theorem 4.1. □

We also use the bar notation to denote the ρ-transformation of the switched linear system $(\Sigma, \mathcal{G})$ in (13) into $(\bar{\Sigma}, \mathcal{G})$ in (21); for each node $r \in \{1, \dots, N_s\}$,

$$G_r = \left(\begin{array}{c|c} \mathcal{A}_r & \mathcal{B}_r \\ \hline \mathcal{C}_r & \mathcal{D}_r \end{array}\right) \quad \text{transforms into} \quad \bar{G}_r = \left(\begin{array}{c|c} \rho^{-1}\mathcal{A}_r & \rho^{-1}\mathcal{B}_r \\ \hline \mathcal{C}_r & \mathcal{D}_r \end{array}\right). \tag{23}$$

Next, we use families of dissipation inequalities satisfied by gradients of functions f in $\mathcal{S}^0_{m,L}$ to construct multiplier relaxations that can certify ρ-convergence of switched algorithms.

Lemma 4.2 (Lemma 5 of [9]). *Consider a function $f^0 \in \mathcal{S}^0_{m,L}$, a rate $\rho \in (0, 1]$, and a coefficient vector $\lambda := (\lambda_0, \dots, \lambda_{\nu_{\max}}) \in \mathbb{R}^{\nu_{\max}+1}$ satisfying*

$$\lambda_\nu \leq 0 \quad \forall \nu \geq 1 \quad \text{and} \quad \textstyle\sum_{\nu=0}^{\nu_{\max}} \rho^{-\nu}\lambda_\nu > 0. \tag{24}$$

For any sequence $\bar{z}$, define the transformed sequences $\bar{q}, \bar{p}, \bar{g}$ as

$$\bar{q}_k := \rho^{-k}\nabla f^0(\rho^k \bar{z}_k) - m\bar{z}_k, \tag{25a}$$

$$\bar{p}_k := L\bar{z}_k - \rho^{-k}\nabla f^0(\rho^k \bar{z}_k), \qquad \bar{g}_k := \textstyle\sum_{\nu=0}^{k} \lambda_\nu \bar{p}_{k-\nu}. \tag{25b}$$

Then the sequences $\bar{q}$ and $\bar{g}$ satisfy the following dissipation inequality:

$$\textstyle\sum_{k=0}^{T-1} \bar{q}_k^\top \bar{g}_k \geq 0 \qquad \forall\, T \in \mathbb{N}, T > 0. \tag{26}$$

The coefficients λ in (24) parameterize a Zames-Falb Finite Impulse Response (FIR) filter with transfer function $\Psi(\lambda) = \sum_{\nu=0}^{\nu_{\max}} \lambda_\nu \mathbf{z}^{-\nu}$ which maps $\bar{p}$ into $\bar{g}$ as in (25). A state space realization of $\Psi(\lambda)$ may be defined using fixed matrices A_f, B_f, and λ-affine maps $C_f(\lambda), D_f(\lambda)$ as

$$\left(\begin{array}{c} \psi_{k+1} \\ \hline \bar{g}_k \end{array}\right) = \left(\begin{array}{c|c} A_f & B_f \\ \hline C_f(\lambda) & D_f(\lambda) \end{array}\right) \left(\begin{array}{c} \psi_k \\ \hline \bar{p}_k \end{array}\right) \quad \text{with} \quad \psi_0 = 0. \tag{27}$$

The choice $\lambda_0 = 1$ and $\lambda_\nu = 0$ for all $\nu \geq 1$ leads to a passivity property (identity filter). Cases with $\lambda_\nu \neq 0$ for some $\nu \geq 1$ describe dynamic filters.

Applying the transformations in Lemma 4.2 to trajectories of (21) gives

$$\begin{pmatrix} \bar{p}_k \\ \bar{q}_k \end{pmatrix} = \begin{pmatrix} LI & -I \\ -mI & I \end{pmatrix} \begin{pmatrix} \bar{z}_k \\ \bar{w}_k \end{pmatrix} \iff \begin{pmatrix} \bar{p}_k \\ \bar{w}_k \end{pmatrix} = \begin{pmatrix} -(L-m)I & I \\ I & mI \end{pmatrix} \begin{pmatrix} \bar{q}_k \\ \bar{z}_k \end{pmatrix}. \tag{28}$$

Based on the given switched system $(\bar{\Sigma}, \mathcal{G})$ with input $\bar{w}$ and output $\bar{z}$, we now construct the switched system $(\hat{\Sigma}^\lambda, \mathcal{G})$ with input $\bar{q}$ and output $\bar{g}$ by

$$\hat{G}_r^\lambda = \Psi(\lambda)\left[\begin{pmatrix} -(L-m)I & I \\ I & mI \end{pmatrix} \star \bar{G}_r\right] \quad \text{described with} \quad \left(\begin{array}{c|c} \hat{\mathcal{A}}_r & \hat{\mathcal{B}}_r \\ \hline \hat{\mathcal{C}}_r^\lambda & \hat{\mathcal{D}}_r^\lambda \end{array}\right) \tag{29}$$

for every $r \in \{1, \ldots, N_s\}$. Note that $\hat{\mathcal{A}}_r$ and $\hat{\mathcal{B}}_r$ are independent from λ, while $\hat{\mathcal{C}}_r^\lambda$ and $\hat{\mathcal{D}}_r^\lambda$ are affine in λ. This leads us to the following analysis result.

Theorem 4.3. *The interconnection (4) is certified as a ρ-exponentially convergent switched optimization algorithm if (12) is satisfied and there exists a filter length $\nu_{\max} \in \mathbb{N}$, filter coefficients $\lambda \in \mathbb{R}^{\nu_{\max}+1}$ satisfying (24), and matrices $M_r \in \mathbb{S}^{n+\nu_{\max}}$, $M_r \succ 0$ for each $r \in 1, \ldots, s$ such that the following LMIs are satisfied for all $(r, r') \in \mathcal{E}$:*

$$[\star]^\top \begin{pmatrix} M_{r'} & 0 \\ 0 & -M_r \end{pmatrix} \begin{pmatrix} \hat{\mathcal{A}}_r^\lambda & \hat{\mathcal{B}}_r^\lambda \\ I & 0 \end{pmatrix} + [\star]^\top \begin{pmatrix} 0 & I_c \\ I_c & 0 \end{pmatrix} \begin{pmatrix} \hat{\mathcal{C}}_r^\lambda & \hat{\mathcal{D}}_r^\lambda \\ 0 & I_c \end{pmatrix} \prec 0. \tag{30}$$

Proof: A detailed proof proceeds as in [9, 31]. The essential arguments are as follows. Feasibility of the LMIs (30) guarantees strict antipassivity of the switched system defined by (29) with the input $\bar{q}$ and the output $\bar{g}$. Lemma 4.2 details that gradients of functions in $\mathcal{S}_{m,L}^0$ induce the passivity property (26) for the signals $\bar{q}$ and $\bar{g}$. A routine passivity argument permits to draw the conclusion that (21) is bounded, which in turn guarantees ρ-exponential convergence for the algorithmic interconnection (4). □

The problem in (30) is a finite-dimensional convex program in M_r and λ for fixed ρ and filter order $\nu_{\max}$. Using an interior point method to verify feasibility of the system of LMIs (30) via semidefinite programming within a numerical tolerance of ϵ involves $O(|\mathcal{E}|^{1/2} c(n+\nu_{\max}) \log(\epsilon^{-1}))$ steps, with a per-step iteration complexity of $O(|\mathcal{E}| c^6 (n+\nu_{\max})^6)$ [32]. This worst-case complexity can be reduced by using solvers specialized for Kalman-Yakubovich-Popov-Lemma problems [33].

For a given $\nu_{\max}$, a best possible bound on the exponential convergence rate can be determined by performing a bisection on ρ. The accuracy of the bound can be increased by increasing $\nu_{\max}$, but there is no guarantee of convergence to the true rate bound.

5. Algorithm Synthesis

We now focus on the synthesis of switching optimization methods. The algorithm synthesis problem with a fixed filter $\Psi(\lambda)$ satisfying (24) and convergence bound ρ reads as follows.

Problem 2. *Given a function class $\mathcal{S}_{m,L}$ and a network $(P, \mathcal{G})$ defined by (7b), synthesize a controller $(K, \mathcal{G})$ defined by (7c) such that the controlled system $(G, \mathcal{G}) = (P \star K, \mathcal{G})$ leads to a ρ-exponentially convergent switched algorithm (4).*

5.1. Assumptions

The regulation property in Theorem 4.1 will be guaranteed through the introduction of an internal model, which is only possible if a suitable regulator equation for the network $(P, \mathcal{G})$ is solvable.

Lemma 5.1. *Suppose (4) with $(G, \mathcal{G}) = (P \star K, \mathcal{G})$ is a convergent switched optimization algorithm and $\{\Theta_r\}_{r=1}^{N_s}$ satisfy (12) for $\{G_r\}_{r=1}^{N_s} = \{P_r \star K_r\}_{r=1}^{N_s}$. Then there exist solutions $\{\Pi_r, \Gamma_r\}_{r=1}^{N_s}$ of the following regulator equation:*

$$\forall (r, r') \in \mathcal{E} : \qquad \begin{pmatrix} A_r & B_{r,2} \\ C_{r,1} & D_{r,12} \end{pmatrix} \begin{pmatrix} \Pi_r \\ \Gamma_r \end{pmatrix} = \begin{pmatrix} \Pi_{r'} \\ I_c \end{pmatrix}. \tag{31}$$

Proof: Fix $(r, r') \in \mathcal{E}$ and partition Θ_r as $\Theta_r = [\Pi_r^\top, \Xi_r^\top]^\top$. Since well-posedness of $P_r \star K_r$ requires that $I - D_2^K D_{r,22}$ is invertible, the matrix

$$\Gamma_r := (I - D_r^K D_{r,22})^{-1} D_r^K (C_{2,r}\Xi_r + C_{2,r}\Pi_r) \tag{32}$$

is well-defined and easily seen to satisfy (31). □

To perform synthesis of ρ-convergent switched optimization algorithms that can be certified by Theorem 4.3, we require the following assumptions.

Assumption 1. *For the given network $(P, \mathcal{G})$ defined by (7b), the regulator equation (31) has a solution $\{\Pi_r, \Gamma_r\}_{r=1}^{N_s}$.*

Assumption 2. *The matrix $I - D_{r,11}m$ is invertible for each $r \in \{1, \ldots, N_s\}$.*

If Assumption 1 is violated, then there cannot exist a controller K such that $P \star K$ defines a convergent switched optimization algorithm certified by Theorem 4.1 (Lemma 5.1). This assumption refers to a property of the network P only, and is independent of K. Assumption 2 ensures that $(mI) \star \bar{P}_r$ is well-posed, which is required to set up the plant for synthesis in Section 5.3.

5.2. Regulation

Given a solution $\{\Pi_r, \Gamma_r\}_{r=1}^{N_s}$ to (31), we define $\Phi_r := C_{r,2}\Pi_r + D_{r,22}\Gamma_r$ for each $r \in \{1, \dots, N_s\}$. We construct controllers K as the interconnection of a fixed internal model Q and a to-be-designed subcontroller R. An switched internal model Q inspired by [19] (originally derived for a non-switched system) is described for each $r \in \{1, \dots, N_s\}$ by

$$Q_r : \qquad \left(\begin{array}{c} \omega_{k+1} \\ \hline u_k \\ \tilde{y}_k \end{array}\right) = \left(\begin{array}{c|c:cc} I_c & 0 & I & 0 \\ \hline -\Gamma_r & 0 & 0 & I \\ \Phi_r & I & 0 & 0 \end{array}\right) \left(\begin{array}{c} \omega_k \\ \hline y_k \\ \hdashline \tilde{u}_{1k} \\ \tilde{u}_{2k} \end{array}\right). \tag{33}$$

The subcontroller R is described for each $r \in \{1, \dots, N_s\}$ by

$$R_r : \qquad \left(\begin{array}{c} \tilde{\xi}_{k+1} \\ \hline \tilde{u}_{1,k} \\ \tilde{u}_{2,k} \end{array}\right) = \left(\begin{array}{c|c} A_{r,c} & B_{r,c} \\ \hline C_{r,c1} & D_{r,c1} \\ C_{r,c2} & D_{r,c2} \end{array}\right) \left(\begin{array}{c} \tilde{\xi}_k \\ \hline \tilde{y}_k \end{array}\right). \tag{34}$$

Then the controller K in (7c) is expressed as the interconnection $K = Q \star R$, where $K_r = Q_r \star R_r$ is given for each $r \in \{1, \dots, N_s\}$ as

$$K_r : \left(\begin{array}{c|c} A_r^K & B_r^K \\ \hline C_r^K & D_r^K \end{array}\right) = \left(\begin{array}{c|c:c} I + D_{r,c1}\Phi_r & C_{r,c1} & D_{r,c1} \\ \hline B_{r,c}\Phi_r & A_{r,c} & B_{r,c} \\ -\Gamma_r + D_{r,c2}\Phi_r & C_{r,c2} & D_{r,c2} \end{array}\right). \tag{35}$$

The subcontroller R is subject to the constraint $D_{r,11} + D_{r,12}D_{r,c2}D_{r,21} = 0$ for all modes $r \in \{1, \dots, N_s\}$, since this ensures that $P_r \star Q_r \star R_r$ admits a state-space representation as in (4) (with $\mathcal{D}_r = 0$) for all $r \in \{1, \dots, N_s\}$.

Theorem 5.2. *If Assumption 1 is satisfied, then the controller structure (35) ensures that $P \star Q \star R$ satisfies condition (12) for output regulation in Theorem 4.1.*

Proof: This follows from standard arguments in regulator theory [18] with respect to the model structure in [19]. In fact, $P_r \star Q_r \star R_r$ is described by

$$\mathcal{A}_r = \left(\begin{array}{cc|c} A_r + B_{r,2}D_{r,c2}C_{r,2} & -B_{r,2}(\Gamma_r - D_{r,c2}\Phi_r) & B_{r,2}C_{r,c2}, \\ D_{r,c1}C_{r,2} & I + D_{r,c1}\Phi_r & C_{r,c1} \\ \hline B_{r,c}C_{r,2} & -B_{r,c}\Phi_r & I - A_{r,c} \end{array}\right),$$

$$\mathcal{B}_r = \begin{pmatrix} B_{r,1} \\ D_{r,c1}D_{r,11} \\ B_{r,c}D_{r,11} \end{pmatrix}, \; \mathcal{C}_r = \begin{pmatrix} (C_{r,1} + D_{r,12}D_{r,c2}C_{r,2})^\top \\ (-D_{r,12}(\Gamma_r - D_{r,c2}\Phi_r))^\top \\ (D_{r,12}C_{r,c2})^\top \end{pmatrix}^\top \text{ and } \; \mathcal{D}_r = 0. \quad (36)$$

Choosing $\Theta_r = [-\Pi_r^\top, I, 0]^\top$ and $\Theta_{r'} = [-\Pi_{r'}^\top, I, 0]^\top$ yields

$$\begin{pmatrix} \mathcal{A}_r \\ \mathcal{C}_r \end{pmatrix} \Theta_r - \Theta_{r'} = \left(\begin{array}{c} \Pi_{r'} - A_r\Pi - B_{r,2}\Gamma_r \\ -D_{r,c1}(C_{r,2}\Pi_r - (C_{r,2}\Pi_r)) \\ \hdashline -B_{r,c1}(C_{r,2}\Pi_r - (C_{r,2}\Pi_r)) \\ \hline -(C_{r,1}\Pi_r + D_{r,12}\Gamma_r) \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \\ \hdashline 0 \\ \hline I_c \end{array}\right), \quad (37)$$

which certifies (12). □

5.3. Controller Synthesis

The controller synthesis problem can be posed as finding a sub-controller $(R, \mathcal{G})$ ensuring that $(P \star Q \star R, \mathcal{G})$ defines a ρ-convergent switched optimization algorithm. For a fixed filter coefficient λ with (24), the control synthesis problem is solved by finding per-mode subcontrollers $\{R_r\}_{r=1}^{N_s}$ such that the system $\hat{G}^\lambda$ with individual modes defined for each $r \in \{1, \ldots, N_s\}$ as

$$\hat{G}_r^\lambda := \Psi(\lambda) \left[\begin{pmatrix} (L-m)I & I \\ I & mI \end{pmatrix} \star \bar{P}_r \star \bar{Q}_r \star \bar{R}_r \right] \quad (38)$$

satisfies the antipassivity LMI in (30) for all arcs $(r, r') \in \mathcal{E}$. The controller synthesis and reconstruction procedure follows the technique in [22] for Linear

Parameter Varying systems. Our overall procedure for controller synthesis is based on an alternating search between λ and R as described in Algorithm 1. After IterMax iterations, Algorithm 1 will return an exponential convergence rate ρ and a subcontroller R such that $(P \star Q \star R, \mathcal{G})$ defines a ρ-convergent switched optimization algorithm.

Algorithm 1 Synthesis of Switched Optimization Algorithms

Require: A switching graph $\mathcal{G}$, a networked system P described by (7b), constants (m, L), filter length $\nu_{\max}$, and number of iterations IterMax

1: iter $\leftarrow 1$
2: Initial identity filter $\Psi(\lambda) \leftarrow 1$
3: Compute internal models Q by (33)
4: **for** iter $\in 1, \ldots,$ IterMax **do**
5: **Synthesis:** $R \leftarrow$ Given $P \star Q$ and a fixed filter parameter λ, infimize ρ over subcontrollers R such that (30) is feasible for all $(r, r') \in \mathcal{E}$
6: **Analysis:** $\lambda \leftarrow$ Given $P \star Q \star R$, infimize ρ over filter coefficients λ such that the conditions in Theorem 4.3 are satisfied
7: **end for**
8: **return** rate ρ, filter λ, subcontroller R

6. Numerical Examples

Numerical experiments are implemented in MATLAB (2024a). All LMIs are solved using LMIlab [34] from the Robust Control Toolbox of MATLAB. The code to generate experiments is publicly available[1]. These examples aim to find the optimal value of the following quadratic plus log-sum-exp function parameterized by a symmetric matrix $\Lambda \in \mathbb{R}^{c \times c}$, $b \in \mathbb{R}^c$, and $\ell \in \mathbb{R}$:

$$f(z; \Lambda, b, \ell) = \frac{1}{2} z^\top \Lambda z + b^\top z + \ell \log \left(\sum_{i=1}^{d} e^{z_i} \right). \tag{39}$$

[1] `https://github.com/jarmill/time_var_opt`

The optimal value of (39) must be found when the oracle $\nabla f(\cdot; \Lambda, b, \ell)$ is only accessible through a time-varying network. Scalars $0 < m < L < \infty$ are given to define the function class (39). Both $b \in \mathbb{R}^c$ and the symmetric matrix Λ are randomly generated and fixed, such that Λ has its eigenvalues in $[m, L']$ for a given $L' \in (m, L)$, while ℓ equals $\ell := L - L'$. Under these choices of (b, Λ, ℓ), every instance of f from (39) is a member of $\mathcal{S}_{m,L}$.

6.1. Inhomogenous Networked System

We first consider minimization of f in (39) over a network governed by $N_s = 4$ switching modes. The switching logic is displayed in Figure 2.

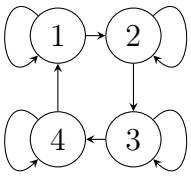


Figure 2: Ring-structured switching graph

The modes $\{P_r\}_{r=1}^4$ of the network P are described by

$$P_1 : \left(\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0.2 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{array}\right) \otimes I_c, \qquad P_2 : \left(\begin{array}{cc|cc} 0.2 & 0 & 0.25 & 0 \\ 0 & 0.9 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 \\ -0.4 & 0 & -0.5 & 3 \end{array}\right) \otimes I_c,$$
$$P_3 : \left(\begin{array}{cc|cc} -0.3 & 0 & 0.5 & 0 \\ 0 & -0.5 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 \\ -0.3 & 0 & 0.5 & 1 \end{array}\right) \otimes I_c, \quad P_4 : \left(\begin{array}{cc|cc} 1.2 & 0 & 1 & 0 \\ 0 & -0.2 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 2 \end{array}\right) \otimes I_c. \tag{40}$$

The systems in (40) are chosen such that unstable channel dynamics are present in the subsystems (in P_4), and that (31) admits a solution with non-

identical $(\Gamma,\ \Phi)$ values between modes, with the specific values of

$$\begin{pmatrix}\Gamma_r\\ \Phi_r\end{pmatrix}_{r=1}^{4} = \begin{pmatrix}-0.8 & -0.1 & -1.5 & -1.2\\ 0 & -0.3 & -1.5 & -2.4\end{pmatrix} \otimes I_c. \tag{41}$$

The Π values are the same between modes, with $\Pi_r = [0,1]^\top \otimes I_c$ for all $r \in \{1,\dots,4\}$. The solutions (Π,Γ,Φ) in (41) are unique. Thus, the method from [26] requiring mode-independent solutions of the regulator equations cannot be used to certify regulation.

We ran Algorithm 1 for 3 iterations. With an identity multiplier ($\Psi(\lambda) = 1$), a switched subcontroller R of order 3 in each mode r is obtained yielding $\rho \leq 0.8944$. With Zames-Falb multipliers of maximal order $\nu_{\max} = 3$, the analysis routine certifies a controller R of order 6 in each mode r with $\rho \leq 0.8913$ and the multiplier $(\lambda_0,\lambda_1,\lambda_2,\lambda_3) = (1,-0.3922,-0.2977,-1.403\times 10^{-6})$. Figure 3 plots traces of optimizing an instance of f over the network with (40) with $L' = 1.25, L = 1.6, c = 400$ by the developed controller. When starting at a randomly drawn initial state, the iterates z converge to the optimum within 30 time steps (top left), and the gradient converges to zero (top right). The switching signal s_k (bottom-left) arises from traversing the graph in Figure 2 in a random walk choosing an edge uniformly at random. The suboptimality in the function value (bottom-right) shows that $f(z_{30}) - f(z^*) \approx 10^{-7}$ after 30 time steps.

We now perform a parameter sweep over increasing values of $L \in (1,10]$ for $m = 1$. The reference algorithms are gradient descent and triple momentum [35] with $\rho_{\text{gd}} = \frac{2}{m+L}$ and $\rho_{\text{TM}} = 1 - \sqrt{\frac{m}{L}}$, which are tuned to optimally converge if there are no network dynamics. These algorithms can be described by

$$K_{\text{gd}}: \left(\begin{array}{c|c} I_c & -\frac{2}{m+L}I_c \\ \hline I_c & 0 \end{array}\right), \quad K_{\text{TM}}: \left(\begin{array}{cc|c} I_c & (1+\frac{\rho_{\text{TM}}^2}{2-\rho_{\text{TM}}})I_c & -\frac{1+\rho_{\text{TM}}}{L}I_c \\ 0 & \frac{\rho_{\text{TM}}^2}{2-\rho_{\text{TM}}}I_c & \frac{1+\rho_{\text{TM}}}{L}I_c \\ \hline I_c & \frac{\rho_{\text{TM}}^2}{(1+\rho_{\text{TM}})(2-\rho_{\text{TM}})}I_c & 0 \end{array}\right). \tag{42}$$

Figure 4 plots an upper-bound on ρ as a function of L obtained by solving the analysis program in Theorem 4.3 with filters of length $\nu_{\max} = 3$. The

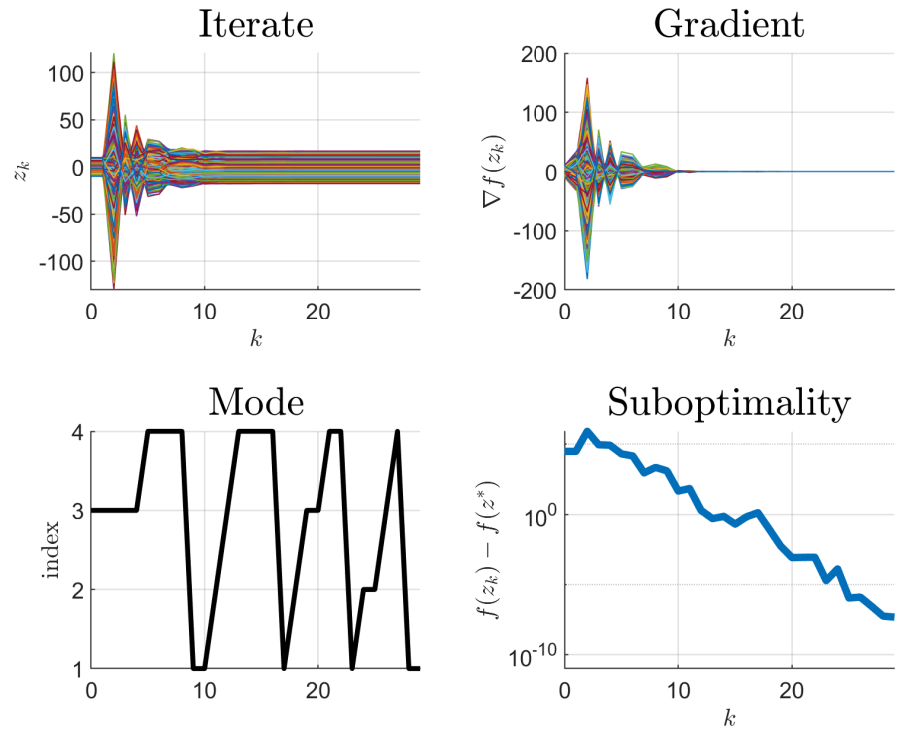


Figure 3: Execution of the designed switched optimization algorithm for $\rho = 0.8913$

compared curves in Figure 4 result from gradient descent, triple momentum, and controllers designed by our procedure with identity filters for two scenarios; one is to let M_r change between different modes $r \in \{1, \ldots, N_s\}$, while the other one involves the constraint that the matrices M_r in the LMI (30) are all identical for $r \in \{1, \ldots, N_s\}$. Setting M_r to the same value among all modes r ignores the ring structure of the switching graph and introduces conservatism. The synthesized switched controller guarantees stability ($\rho < 1$) for $L < 3.360$, while the one obtained for common matrices M_r does so only for $L < 1.660$. Neither gradient decent nor triple momentum can be certified to be stable by analysis; instability can also be empirically confirmed even for the small values $L = 1.01$ and $L' = 1.005$.

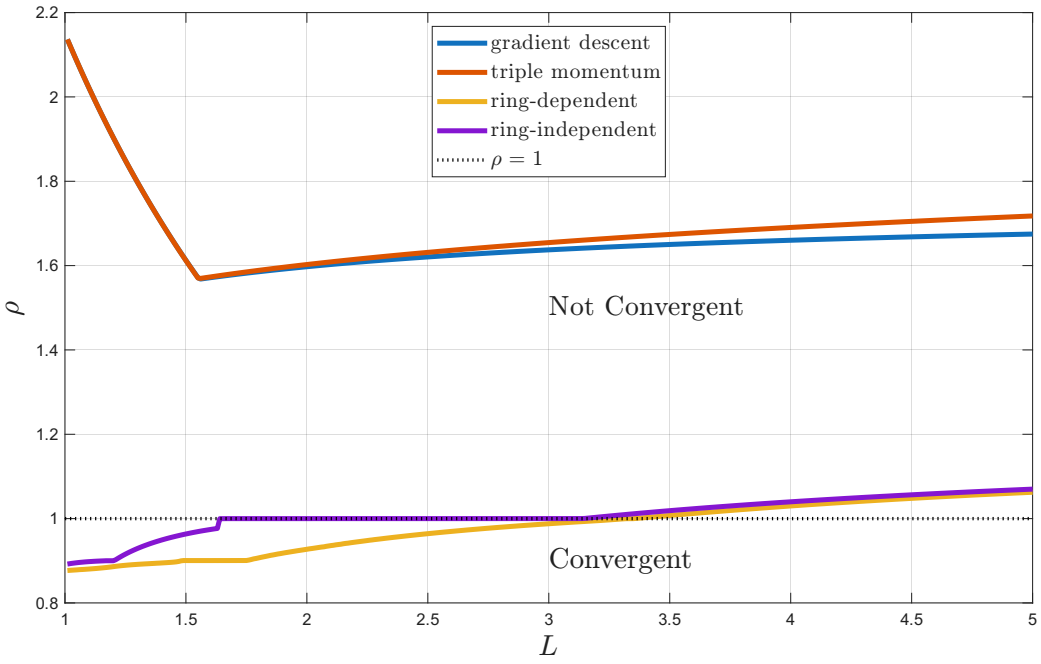


Figure 4: Upper bound on ρ as a function of L by analysis with $\nu_{\max} = 3$ for $m = 1$

6.2. Time-Varying Delay: Increasing Delay

Our next example involves synthesis of algorithms with switching delays before the oracle ∇f, as modeled by the network described in Section 3.2 and with $H_{\max} \in \{1, \dots, 5\}$. We consider the case of unrestricted delays and packet drops. In the unrestricted delay setting, no constraint is imposed on how $h_k \in \{0, \dots, H_{\max}\}$ switches to $h_{k+1} \in \{0, \dots, H_{\max}\}$. The packet drop setting involves the delay increasing by 1 until reaching the maximum value $H_{\max}$, or jumping back to 0. This is visualized in Figure 1b for $H_{\max} = 3$.

Figure 5 plots the computed convergence rates ρ by synthesis under identity multipliers for $m = 1$ as L is swept in the range $(1, 10]$. In each subplot, the color indicates the chosen maximal delay $H_{\max}$. The top subplot shows the computed bounds ρ_{all} for unrestricted delays, while the middle subplot depicts the bounds ρ_{drop} under the packet drop logic, in which the mode/delay at time k depends on the one at time $k - 1$. All bounds on the rate rise as the maximal delay increases. The bottom subplot reveals that the differences

$\rho_{\text{all}} - \rho_{\text{drop}}$ are positive for $H_{\max} \in \{2, 3, 4, 5\}$. This demonstrate that the constraints on switching the delays for the specific packet drop logic generally yields smaller bounds on the convergence rates if compared to unrestricted delay switchings, while the bounds are identical for $H_{\max} \in \{0, 1\}$.

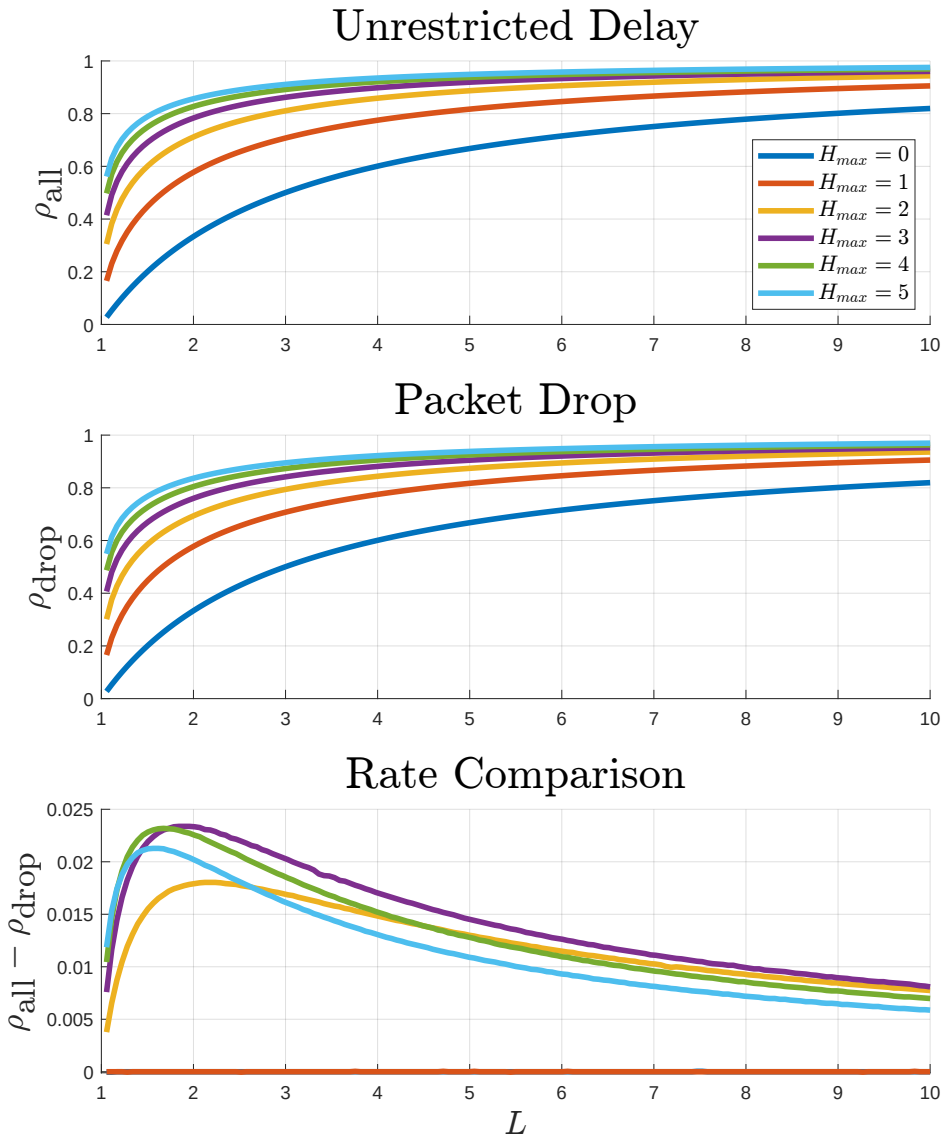


Figure 5: Exponential rates under increasing delay and L for $m = 1$

7. Conclusion

We presented a method to analyze and synthesize switched optimization algorithms. Regulation theory of switched systems is used to certify the convergence of optimization algorithms. Sufficient conditions for the algorithm to converge to the global optimal solution are if the switched system interconnected with a error-centered gradient oracle achieves robust stability, and if a Regulator Equation is feasible (Theorem 4.1). Analysis of exponential convergence rates occurs through convex searches over Zames-Falb filter coefficients λ. Synthesis of switched algorithms given a network P and fixed filters $\Psi(\lambda)$ is accomplished by using an output feedback controller with a switched internal model. The analysis and synthesis tasks are jointly convex, and an alternating approach can be used to develop certified switching algorithms with exponential convergence guarantees. Future work involves the establishing of necessary conditions for convergence of switched optimization algorithms, finding delay-scheduled controllers in the sampled data case, and developing certified optimization algorithms in the unknown delay case. Other work involves formulating convex joint searches over controllers R_r and filter coefficients λ, and extending our method to synthesis of distributed optimization algorithms.

Acknowledgements

J. Miller and C. Scherer are by supported the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 – 390740016. We acknowledge the support by the Stuttgart Center for Simulation Science (SimTech). F. Jakob acknowledges the support of the International Max Planck Research School for Intelligent Systems (IMPRS-IS).

References

[1] L. Lessard, B. Recht, A. Packard, Analysis and design of optimization algorithms via integral quadratic constraints, SIAM J. Optim. 26 (2016) 57–95.

[2] A. I. Lur'e, V. N. Postnikov, On the theory of stability of control systems, Applied mathematics and mechanics 8 (1944) 246–248.

[3] A. Megretski, A. Rantzer, System analysis via Integral Quadratic Constraints, IEEE Transactions on Automatic Control 42 (1997) 819–830.

[4] R. Boczar, L. Lessard, B. Recht, Exponential Convergence Bounds Using Integral Quadratic Constraints, in: American Control Conference, 2018.

[5] C. Scherer, C. Ebenbauer, Convex synthesis of accelerated gradient algorithms, SIAM J. Control Optim. 59 (2021) 4615–4645.

[6] S. Michalowsky, C. Scherer, C. Ebenbauer, Robust and structure exploiting optimisation algorithms: an integral quadratic constraint approach, International Journal of Control 94 (2021) 2956–2979.

[7] H. Mohammadi, M. Razaviyayn, M. R. Jovanović, Robustness of Accelerated First-Order Algorithms for Strongly Convex Optimization Problems, IEEE Transactions on Automatic Control (2021).

[8] T. Holicki, C. W. Scherer, Algorithm design and extremum control: Convex synthesis due to plant multiplier commutation, in: Proc. IEEE Conf. Decis. Control, IEEE, 2021, pp. 3249–3256.

[9] C. W. Scherer, C. Ebenbauer, T. Holicki, Optimization algorithm synthesis based on integral quadratic constraints: A tutorial, in: Proc. IEEE Conf. Decis. Control, 2023, pp. 2995–3002.

[10] F. Jakob, A. Iannelli, A Linear Parameter-Varying Framework for the Analysis of Time-Varying Optimization Algorithms, arXiv:2501.07461 (2025).

[11] Z. Wen'an, Y. Li, S. Hongbo, A switched system approach to networked control systems with time-varying delays, in: 2008 27th Chinese Control Conference, 2008, pp. 424–427.

[12] G. Conte, A. M. Perdon, E. Zattoni, Modeling discrete time systems with variable delays as switching systems without delays, in: 2020 European Control Conference (ECC), IEEE, 2020, pp. 1591–1594.

[13] A. Ramaswamy, A. Redder, D. E. Quevedo, Optimization over time-varying networks and unbounded information delays, IEEE Transactions on Automatic Control 67 (2021) 4131–4137.

[14] M. Doostmohammadian, U. A. Khan, M. Pirani, T. Charalambous, Consensus-based distributed estimation in the presence of heterogeneous, time-invariant delays, IEEE Control Syst. Lett. 6 (2021) 1598–1603.

[15] C. Briat, O. Sename, J.-F. Lafay, Delay-scheduled state-feedback design for time-delay systems with time-varying delays–a LPV approach, Systems & Control Letters 58 (2009) 664–671.

[16] P. Apkarian, P. Gahinet, G. Becker, Self-scheduled H-infinity control of linear parameter-varying systems: a design example, Automatica 31 (1995) 1251–1261.

[17] C. W. Scherer, Robust Exponential Stability and Invariance Guarantees with General Dynamic O'Shea-Zames-Falb Multipliers, IFAC-PapersOnLine 56 (2023) 5799–5804.

[18] B. A. Francis, W. M. Wonham, The internal model principle of control theory, Automatica 12 (1976) 457–465.

[19] A. A. Stoorvogel, A. Saberi, P. Sannuti, Performance with regulation constraints, Automatica 36 (2000) 1443–1456.

[20] H. Köroğlu, C. W. Scherer, LPV control for robust attenuation of non-stationary sinusoidal disturbances with measurable frequencies, IFAC Proceedings Volumes 41 (2008) 4928–4933.

[21] G. Conte, A. Perdon, E. Zattoni, D. Animobono, Disturbance decoupling and model matching problems for discrete-time systems with time-varying delays, Nonlinear Analysis: Hybrid Systems 41 (2021) 101043.

[22] J. De Caigny, J. F. Camino, R. C. Oliveira, P. L. Peres, J. Swevers, Gain-scheduled dynamic output feedback control for discrete-time LPV systems, Int. J. Robust Nonlinear Control. 22 (2012) 535–558.

[23] K. Zhou, J. C. Doyle, Essentials of Robust Control, volume 104, Prentice hall Upper Saddle River, NJ, 1998.

[24] T. Rotaru, F. Glineur, P. Patrinos, Exact worst-case convergence rates of gradient descent: a complete analysis for all constant stepsizes over non-convex and convex functions, arXiv preprint arXiv:2406.17506 (2024).

[25] J. Daafouz, P. Riedinger, C. Iung, Stability analysis and control synthesis for switched systems: a switched Lyapunov function approach, IEEE Transactions on Automatic Control 47 (2002) 1883–1887.

[26] E. Zattoni, A. M. Perdon, G. Conte, The output regulation problem with stability for linear switching systems: A geometric approach, Automatica 49 (2013) 2953–2962.

[27] Y. Zhao, Y. Liu, D. Ma, Output regulation for switched systems with multiple disturbances, IEEE Transactions on Circuits and Systems I: Regular Papers 67 (2020) 5326–5335.

[28] J.-W. Lee, G. E. Dullerud, Uniform stabilization of discrete-time switched and markovian jump linear systems, Automatica 42 (2006) 205–218.

[29] A. A. Ahmadi, R. M. Jungers, P. A. Parrilo, M. Roozbehani, Joint Spectral Radius and Path-Complete Graph Lyapunov Functions, SIAM J. Control Optim. 52 (2014) 687–717.

[30] D. Lee, G. E. Dullerud, J. Hu, Graph Lyapunov function for switching stabilization and distributed computation, Automatica 116 (2020) 108923.

[31] C. W. Scherer, C. Ebenbauer, A Tutorial on Convex Design of Optimization Algorithms by Integral Quadratic Constraints, Annual Review of Control, Robotics, and Autonomous Systems 9 (2025) 215–242.

[32] A. Ben-Tal, A. Nemirovski, Lectures on modern convex optimization: analysis, algorithms, and engineering applications, SIAM, 2001.

[33] L. Vandenberghe, V. R. Balakrishnan, R. Wallin, A. Hansson, T. Roh, Interior-point algorithms for semidefinite programming problems derived from the KYP lemma, in: Positive Polynomials in Control, Springer, 2005, pp. 195–238.

[34] P. Gahinet, A. Nemirovskii, General-Purpose LMI solvers with benchmarks, in: Proc. IEEE Conf. Decis. Control, 1993, pp. 3162–3165 vol.4.

[35] B. Van Scoy, R. A. Freeman, K. M. Lynch, The Fastest Known Globally Convergent First-Order Method for Minimizing Strongly Convex Functions, IEEE Control Systems Letters 2 (2017) 49–54.