

Convex Synthesis of First-Order Methods for Time-Varying Smooth Strongly Convex Optimization

Bryan Van Scoy

Gianluca Bianchin

Abstract—Time-varying optimization is fundamental to decision-making in dynamic environments, where objectives evolve over time due to exogenous signals or data streams. However, algorithms designed for static problems yield suboptimal decisions in dynamic scenarios, even asymptotically. In this paper, we develop a robust control synthesis framework to systematically design first-order methods for smooth strongly convex problems that vary in time. Our approach leverages both convex robust control synthesis in the static setting and the internal model principle by directly embedding a model of the underlying variability into the designed algorithm.

I. INTRODUCTION

Optimization algorithms play a central role in modern control, machine learning, and networked systems, underpinning decision-making processes in applications ranging from large-scale data analytics to cyber-physical and energy systems. Recent advances have established a systems-and-control perspective for algorithmic optimization analysis and design, whereby iterative optimization methods are modeled as dynamical systems in feedback with nonlinear gradient oracles, enabling their analysis and design via tools from dissipativity theory, passivity, and robust control [1], [2]. Building on this viewpoint, convex synthesis approaches [3] have emerged as a principled framework for the automated design of algorithms, casting algorithm construction as a robust control synthesis problem, enabling the systematic exploration of algorithmic structures beyond classical hand-crafted schemes and leading to provably optimal methods [4].

Despite these advances, existing synthesis approaches address *static* optimization problems, while many applications of practical interest involve *time-varying* objectives driven by exogenous signals, evolving environments, or streaming data. In such settings, standard first-order methods (e.g., gradient descent) exhibit nonzero steady-state error [5], as they lack mechanisms to compensate for persistent temporal variations. From a control-theoretic perspective, this limitation stems from the absence of appropriate internal models [6], [7]. More critically, accelerated methods designed for time-invariant settings can suffer a noticeable loss of robustness in time-varying regimes, leading to amplified tracking errors

This material is based upon work supported in part by the National Science Foundation under Award No. 2347121 and in part by the FRFS WEL-T Investigator Programme. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

B. Van Scoy is with the Dept. of Electrical and Computer Engineering, Miami University, OH 45056, USA. Email: bvanscoy@miamioh.edu

G. Bianchin is with the ICTEAM institute and the Department of Mathematical Engineering (INMA) at the University of Louvain, Belgium. Email: gianluca.bianchin@uclouvain.be

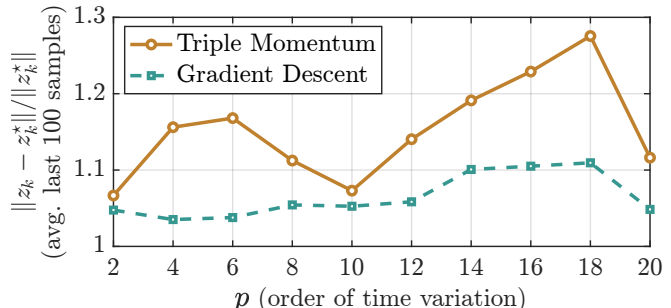


Fig. 1. Asymptotic (averaged over the last 100 iterations) relative tracking error versus the order p of the temporal variation, comparing gradient descent and the accelerated triple momentum method [8] on the time-varying quadratic problem $f(z, \theta_k) = z^T Q z + \theta_k^T z$, with Q random (eigenvalues in $[1, 50]$) and $\theta_{k+1} = S \theta_k$, $S \in \mathbb{R}^{p \times p}$. As p increases, the performance of the accelerated method degrades relative to gradient descent, highlighting the need for acceleration schemes that explicitly account for temporal variability.

and, in some cases, inferior asymptotic performance compared to non-accelerated schemes, as illustrated in Figure 1.

According to the internal model principle of time-varying optimization [6], [7], perfect tracking and disturbance rejection require embedding a model of the exogenous signals within the optimization dynamics. Yet, this principle has not been systematically integrated in convex synthesis [3].

Contributions: Our main contribution is to demonstrate how the convex synthesis procedure from [9] can be adapted to incorporate an internal model of the temporal variability of the problem, thereby extending the approach to be applicable to time-varying optimization. Our design is based on the internal model principle of time-varying optimization [6], [7]. The effectiveness of the approach is illustrated through numerical simulations.

Related works: The systems-and-control perspective on optimization interprets first-order methods as dynamical systems in feedback with gradient oracles, thereby enabling analysis through tools from robust control, dissipativity, and integral quadratic constraints (IQCs) [1], [10]–[12]. This viewpoint has also motivated algorithm synthesis methods that go beyond classical hand-crafted designs, including direct synthesis with worst-case convergence guarantees [13], IQC-based structure-exploiting design [14], linear matrix inequality (LMI)-based synthesis for optimization and saddle-point problems [15], and convex synthesis frameworks for accelerated algorithms and extremum control [3], [16]. In parallel, time-varying optimization has been studied from both online and control-theoretic perspectives. When no

explicit model of temporal variability is available, one typically obtains dynamic-regret or input-to-state-stability type guarantees for online gradient, primal-dual, and related running methods [17]–[20]. When the temporal variability can instead be modeled or estimated, prediction-correction and related model-based schemes can achieve substantially improved tracking performance [21]–[25]. More recently, these directions have started to intersect with robust-control tools for time-varying optimization [7], [23]. At a more fundamental level, the internal-model viewpoint connects exact asymptotic tracking in time-varying optimization to output regulation theory, showing that exact tracking requires embedding a suitable model of the temporal variability within the algorithm itself [7], [24], [26], [27]. Particularly relevant to the design methodology pursued in this work are [28], [29], where it is observed that output regulation of nonlinear models may be achieved by linear controllers, provided that they incorporate more dynamics than those generated by the exosystem. These developments motivate the present work: rather than treating convex synthesis and internal-model design as separate threads, we integrate them into a unified synthesis framework to design optimization algorithms tailored to structured time-varying problems.

Notation: We denote the natural numbers including zero by \mathbb{N} , and the reals by \mathbb{R} . The set of $n \times n$ symmetric matrices is \mathbb{S}^n . We denote the state-space realization (A, B, C, D) of a discrete-time LTI system G as

$$G = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right].$$

II. PROBLEM SETUP

Consider an unconstrained optimization problem in which the objective function depends on a time-varying parameter:

$$\text{minimize } f(z, \theta_k), \quad (1)$$

where $f : \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}$ is the objective function, $z \in \mathbb{R}^d$ the decision variable, and $\theta : \mathbb{N} \rightarrow \mathbb{R}^p$ a time-varying parameter. We make the following assumption on the objective function.

Assumption 1 (Objective): For any fixed parameter $\theta_o \in \mathbb{R}^p$, the function $z \mapsto f(z, \theta_o)$ is L -smooth and μ -strongly convex with known parameters $0 < \mu < L$. \square

The objective being L -smooth means that the gradient is Lipschitz continuous with parameter L , and the objective being μ -strongly convex means that $z \mapsto f(z, \theta_o) - \frac{\mu}{2} \|z\|^2$ is convex. With this assumption, the optimization problem has a unique optimal trajectory $z^* : \mathbb{N} \rightarrow \mathbb{R}^d$. We use $\mathcal{F}_{\mu, L}$ to denote the set of all functions that satisfy Assumption 1.

The optimization problem varies in time due to the parameter θ_k , which we assume is generated by a (known) LTI system.

Assumption 2 (Exosystem): The parameter signal $\theta : \mathbb{N} \rightarrow \mathbb{R}^p$ is generated by an autonomous LTI system satisfying

$$\theta_{k+1} = S\theta_k \quad \text{for all } k \in \mathbb{N} \quad (2)$$

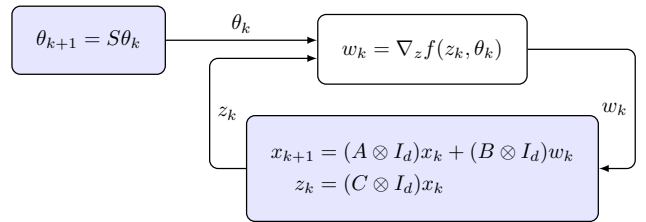


Fig. 2. Block diagram of a first-order method for time-varying optimization. The exosystem generates the time-varying parameter θ_k , which affects the gradient of the objective. The method is a dynamical system with state x_k that generates a sequence of points z_k at which to evaluate the gradient based on measurements w_k of the time-varying gradient.

and some initial condition $\theta_0 \in \mathbb{R}^p$. Moreover, all the eigenvalues of S have unit modulus, all exosystem trajectories are bounded, and the model S is known. \square

The requirement that (2) is known is motivated by [6], [7], which show that knowledge of the exosystem is necessary to compute an exact optimizer of (1). In its absence, one must incur nonzero regret [23]; here, we focus on algorithms that achieve zero regret, and hence impose this requirement.

In this work, we are interested in designing first-order optimization methods that can query the gradient of the objective at each iteration. Specifically, the method selects a point z_k at which to query the gradient, and then receives the corresponding value of the gradient

$$w_k = \nabla_z f(z_k, \theta_k), \quad (3)$$

with $\nabla_z f$ denoting the gradient of f with respect to its first argument. To exclude exosystem modes that do not influence the gradient (and are therefore irrelevant to the optimization), we assume the system contains no nontrivial zero dynamics.

Assumption 3 (Absence of zero dynamics): The objective f and exosystem trajectory θ are such that $\nabla_z f(z, \theta_k) = 0$ for all $k \in \mathbb{N}$ and some $z \in \mathbb{R}^d$ implies $\theta_k = 0$ for all $k \in \mathbb{N}$. \square

An optimization algorithm then consists of a mapping from w to z . For implementability, we require this mapping to be causal, linear, time-invariant, strictly proper, and to act homogeneously across each dimension of \mathbb{R}^d . Formally, we consider methods of the form

$$x_{k+1} = (A \otimes I_d)x_k + (B \otimes I_d)w_k, \quad (4a)$$

$$z_k = (C \otimes I_d)x_k, \quad (4b)$$

for some state-space matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times 1}$, and $C \in \mathbb{R}^{1 \times n}$, where $x_k \in \mathbb{R}^{nd}$ is the state. In line with [30, §2.2], we assume that the pair (A, C) is detectable without loss of generality. The algorithm is illustrated in Figure 2.

For the algorithm (4) to solve time-varying optimization problems, it must in particular be capable of solving time-invariant problems (e.g., when the objective does not depend on θ_k). A necessary condition for this is that the pair (A, C) has an observable eigenvalue at one [9, § II.C]. This, in turn, implies the existence of a fixed point.

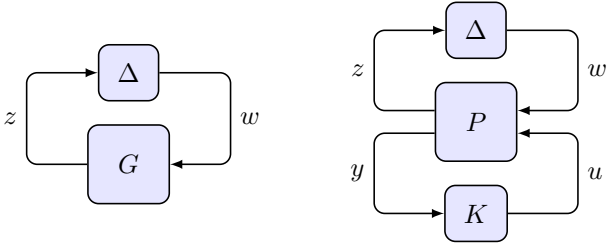


Fig. 3. Block diagrams of system interconnections. (Left) The algorithm transfer function $G(z)$ in feedback with an uncertainty Δ as in Thm. 1. (Right) Plant P in feedback with controller K and uncertainty Δ as in (9).

Assumption 4 (Fixed point): There exist $x_* \in \mathbb{R}^{nd}$ such that

$$x_* = (A \otimes I_d)x_* \quad \text{and} \quad (C \otimes I_d)x_* \neq 0. \quad \square$$

Definition 1: We say that algorithm (4) (locally) asymptotically tracks the optimizer of (1) with rate $\rho \in (0, 1)$ if there exist constants $c > 0$ and $\delta > 0$ such that, for any initial condition x_0 with $\|x_0 - x_*\| < \delta$,

$$\|z_k - z_k^*\| \leq c\rho^k \|x_0 - x_*\| \quad \text{for all } k \in \mathbb{N}, \quad (5)$$

where $z_k^* \in \mathbb{R}^d$ is the unique optimizer at each k and x_k^* is the corresponding fixed point of (4). Moreover, we say that the tracking is *robust with respect to a set of objective functions and exosystem trajectories* if the bound holds for all such instances.

With this setup, we can now state our main goal.

Problem 1: Design an efficient procedure that, given the objective function parameters μ and L and the temporal variability model S , automatically selects the state-space matrices (A, B, C) in (4) such that the resulting method robustly asymptotically tracks the optimizer of (1) with optimal (i.e., minimal) rate ρ . Optimality is with respect to all admissible choices of state-space matrices (A, B, C) , and robustness is with respect to all objective functions f and exosystem trajectories θ that satisfy Assumptions 1 to 3. \square

III. INTERNAL MODEL EMBEDDING

In this section, we derive a structure for (4) to ensure that the algorithm achieves asymptotic tracking. This will then enable us to reformulate Problem 1 as a robust control problem.

Achieving asymptotic tracking of the optimizer imposes specific structural requirements on the method. In the time-invariant case (in which the model of the time variation is $S = I$, so that θ_k is constant), it is well-known that the method (4) must have the form of an LTI system in series with an integrator [30, §2.2]. When the objective varies in time, by the *internal model principle of time-varying optimization* [6], [7], this concept generalizes to the requirement that the method must embed knowledge of the exosystem in its dynamics.

Building on this finding, we will now present the main result of this work. To state our result, we let Δ denote the set of static maps Δ such that $w = \Delta(z)$ if and only if $w_k = \nabla f(z_k)$ for some function $f \in \mathcal{F}_{\mu, L}$ with $\nabla f(0) = 0$.

Moreover, given a multi-index $\alpha \in \mathbb{N}^p$ and eigenvalues $\lambda = (\lambda_1, \dots, \lambda_p)$ of S , we define $\lambda^\alpha := \lambda_1^{\alpha_1} \lambda_2^{\alpha_2} \dots \lambda_p^{\alpha_p}$. We then define the set of *harmonics* associated with λ as

$$\Omega_\lambda := \{\lambda^\alpha \mid \alpha \in \mathbb{N}^p\}. \quad (6)$$

That is, Ω_λ contains all possible products of the eigenvalues of S , including repeated products, and characterizes all modes that can arise from their combinations.

Example 1: If the exosystem contains a single frequency $\theta = \pi/3$, then S has eigenvalues $\lambda = (e^{j\pi/3}, e^{-j\pi/3})$, in which case the set of harmonics is $\Omega_\lambda = \{0, \frac{\pi}{3}, \frac{2\pi}{3}, \pi\}$.

Theorem 1 (Internal model-based optimization design): Let $0 < \mu < L$ and $\rho \in (0, 1)$. Let (A, B, C) be the state-space matrices of an algorithm (4). Then, the following statements are equivalent:

- (i) Algorithm (4) locally asymptotically tracks the optimizer of (1) with rate ρ for all functions f and exosystem trajectories θ that satisfy Assumptions 1 to 3.
- (ii) There exists a proper rational transfer function $K(z)$ such that $G(z) = C(zI - A)^{-1}B$ factors as

$$G(z) = K(z)H(z) \quad \text{where} \quad H(z) := \frac{z^{|\Omega_\lambda|-1}}{\prod_{\omega \in \Omega_\lambda} (z - \omega)}, \quad (7)$$

and the feedback interconnection of algorithm (4) with uncertainty Δ (see Figure 3) is robustly exponentially stable¹ with rate ρ for all $\Delta \in \Delta$. \square

The proof of this claim is postponed to the appendix.

It follows from Theorem 1 that designing an algorithm (4) to track the optimizer of the time-varying optimization problem (1) is equivalent to solving a robust control problem with plant P and uncertainty set Δ . From a control-theoretic viewpoint, the structural condition (7) admits the following interpretation: the transfer function $H(z)$ captures the required internal model through its pole structure and is thus regarded as the *plant*. In turn, $K(z)$ shapes the behavior of $H(z)$ and is therefore interpreted as the *controller*.

Let a state-space realization of the plant be

$$H = \left[\begin{array}{c|c} A_p & B_p \\ \hline C_p & 0 \end{array} \right]. \quad (8)$$

By Theorem 1, the algorithm must then have the form shown in Figure 3 (right) with plant

$$\begin{bmatrix} z \\ y \end{bmatrix} = P \begin{bmatrix} w \\ u \end{bmatrix} \quad \text{where} \quad P = \left[\begin{array}{c|cc} A_p & B_p & 0 \\ \hline 0 & 0 & 1 \\ C_p & 0 & 0 \end{array} \right]. \quad (9)$$

Let $n_p := |\Omega_\lambda|$ denote the dimension of the plant, which is the number of exosystem harmonics.

¹Robust exponential stability with rate $\rho \in (0, 1)$ amounts to ensuring that there exists $c > 0$ such that, for all $\Delta \in \Delta$ and all x_0 , the closed-loop state x_k satisfies $\|x_k\| \leq c\rho^k \|x_0\|$ for all $k \geq 0$.

IV. ROBUST CONTROL–BASED DESIGN OF METHODS FOR TIME-VARYING OPTIMIZATION

With Problem 1 reformulated as a robust control problem, we now propose the use of the convex synthesis methodology from [9], [31] to design the controller $K(z)$. Once the controller is obtained, the algorithm transfer function is given in (7), the state-space matrices (A, B, C) are any realization of this transfer function, and the algorithm is (4).

A. Transformations

Following this strategy, we first apply a chain of transformations to put the system into standard form [4]. As a first transformation, we exponentially weight the signals so that stability of the transformed system is equivalent to exponential stability of the original system. To that end, define the exponential signal weighting map

$$T_\rho(z_0, z_1, z_2, \dots) = (z_0, \rho z_1, \rho^2 z_2, \dots)$$

with parameter $\rho > 0$. The map T_ρ is linear and invertible with inverse $T_{\rho^{-1}}$. The transformed system is shown in Figure 4, where the transformed plant is

$$\begin{bmatrix} \bar{z} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} \rho^{-1}A_p & \rho^{-1}B_p & 0 \\ 0 & 0 & 1 \\ C_p & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{w} \\ \bar{u} \end{bmatrix}. \quad (10)$$

The transformed system is Lyapunov stable² if and only if the original system is exponentially stable with rate ρ [4].

We next filter the input and output of the (transformed) uncertainty through a static filter to produce the signals \bar{p} and \bar{q} . The system from (\bar{q}, \bar{u}) to (\bar{p}, \bar{y}) is

$$\begin{bmatrix} \bar{p} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} \rho^{-1}A_p & \rho^{-1}B_p & \rho^{-1}\mu B_p \\ 0 & -1 & L - \mu \\ C_p & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{q} \\ \bar{u} \end{bmatrix}.$$

All admissible trajectories of the transformed system satisfy the passivity property $0 \leq \sum_{k=0}^T \bar{q}_k^\top \bar{p}_k$ for all $T \in \mathbb{N}$ [4]. In fact, we can generate a parameterized set of constraints by filtering \bar{p} through the following system, known as a Zames–Falb multiplier [32],

$$\psi = \begin{bmatrix} A_f & B_f \\ C_f & D_f \end{bmatrix} = \begin{bmatrix} 0 & 1 & \dots & 0 & | & 0 \\ 0 & 0 & \ddots & \vdots & | & \vdots \\ \vdots & \vdots & \ddots & 1 & | & 0 \\ 0 & 0 & \dots & 0 & | & 1 \\ \hline \lambda_\ell & \lambda_{\ell-1} & \dots & \lambda_1 & | & \lambda_0 \end{bmatrix}.$$

The output \bar{r} of this filter also satisfies the passivity property $0 \leq \sum_{k=0}^T \bar{q}_k^\top \bar{r}_k$ for all $T \in \mathbb{N}$ if the filter parameters $\lambda_0, \dots, \lambda_\ell \in \mathbb{R}$ belong to the set

$$\Lambda_\ell^\rho = \left\{ (\lambda_0, \dots, \lambda_\ell) \mid \lambda_1, \dots, \lambda_\ell \leq 0, \sum_{j=0}^{\ell} \rho^{-j} \lambda_j \geq 0 \right\}.$$

²Exponential stability means there exists $c > 0$ such that, for every \bar{x}_0 , $\|\bar{x}_k\| \leq c \rho^k \|\bar{x}_0\|$ for all k . Lyapunov stability is the same with $\rho = 1$.

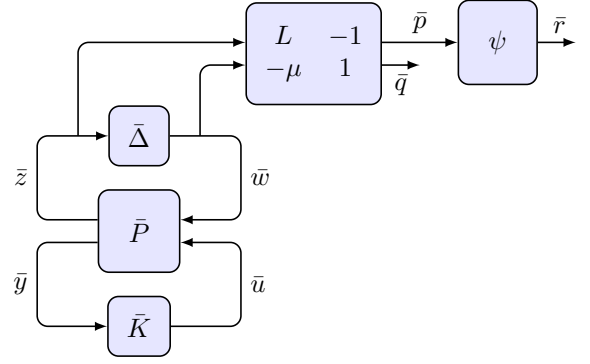


Fig. 4. Equivalent robust control setup after exponential scaling and filtering. Signals are scaled by $T_{\rho^{-1}}$, e.g., $\bar{u} = T_{\rho^{-1}}u$, and the plant is transformed as $\bar{P} = T_{\rho^{-1}} \circ P \circ T_\rho$, and similarly for the uncertainty and controller. The input \bar{z} and output \bar{w} of the transformed uncertainty $\bar{\Delta}$ are then passed through a static map, with the first output \bar{p} then passed through a dynamic filter ψ resulting in \bar{r} .

The transformed plant combined with the filter yield

$$\begin{bmatrix} \bar{r} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} A_f & 0 & -B_f & (L - \mu)B_f \\ 0 & \rho^{-1}A_p & \rho^{-1}B_p & \rho^{-1}\mu B_p \\ C_f & 0 & -D_f & (L - \mu)D_f \\ 0 & C_p & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{q} \\ \bar{u} \end{bmatrix}.$$

For notational convenience, we denote this system by

$$\hat{P} = \begin{bmatrix} \hat{A} & \hat{B}_w & \hat{B} \\ \hat{C}_z & \hat{D}_{zw} & \hat{D}_z \\ \hat{C} & 0 & 0 \end{bmatrix}.$$

To summarize, for any transformed uncertainty $\bar{\Delta}$ with $\Delta \in \mathbf{\Delta}$, all trajectories of the transformed system in Figure 4 satisfy the passivity property $0 \leq \sum_{k=0}^T \bar{q}_k^\top \bar{r}_k$ for all $T \in \mathbb{N}$ whenever the filter parameters satisfy $\lambda \in \Lambda_\ell^\rho$.

B. Analysis

Before designing controllers, we begin with the simpler problem of analyzing the convergence rate of a fixed controller. Let K be a controller of dimension n_c , and let $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ denote a state-space realization of the corresponding system from \bar{q} to \bar{r} in Figure 4. We then have the following result.

Proposition 1 (Analysis): If there exists a matrix $\mathcal{X} \in \mathbb{S}^{n_p + \ell + n_c}$ and filter parameters $\lambda \in \Lambda_\ell^\rho$ such that

$$\mathcal{X} \succ 0, \quad \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \\ 0 & 1 \end{pmatrix}^\top \begin{pmatrix} \mathcal{X} & 0 & 0 & 0 \\ 0 & -\mathcal{X} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \\ 0 & 1 \end{pmatrix} \prec 0, \quad (11)$$

then the interconnection in Figure 3 (right) is robustly exponentially stable with rate ρ for all uncertainties $\Delta \in \mathbf{\Delta}$. \square

C. Convex synthesis with fixed multiplier

While the analysis condition enables us to evaluate the rate of a given controller, our goal is to design controllers. Simply searching over controller matrices is not tractable, as the

analysis conditions are not simultaneously convex in \mathcal{X} , Λ_ℓ^ρ , and the state-space matrices of K . When the multiplier ψ is fixed, however, the controller synthesis follows from seminal \mathcal{H}_∞ synthesis results [33]. To that end, let

$$\hat{U} = \text{diag}(I_\ell, C_{p\perp}, 1), \quad \hat{V}^\top = \begin{bmatrix} B_f^\top & \frac{\mu}{\rho(L-\mu)} B_p^\top & D_f^\top \end{bmatrix}_\perp$$

where M_\perp denotes a matrix whose columns are a basis for the kernel of M . Then, there exists a controller K and matrix $\mathcal{X} \succ 0$ such that the analysis LMI (11) is feasible if and only if there exist symmetric matrices $\hat{X}, \hat{Y} \in \mathbb{S}^{n_p+\ell}$ such that

$$[\cdot]^\top \begin{pmatrix} \hat{X} & 0 & 0 & 0 \\ 0 & -\hat{X} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{bmatrix} \left(\begin{array}{c|c} \hat{A} & \hat{B}_w \\ \hline \hat{C}_z & \hat{D}_{zw} \end{array} \right) \\ \left(\begin{array}{c|c} 0 & 1 \end{array} \right) \end{bmatrix} \hat{U} \prec 0,$$

$$\begin{bmatrix} \hat{V} \left(\begin{array}{c|c} -I & \hat{A} \\ \hline 0 & \hat{C}_z \end{array} \middle| \begin{array}{c} 0 \\ \hat{B}_w \end{array} \right) \\ \left(\begin{array}{c|c} 0 & \hat{D}_{zw} \end{array} \right) \end{bmatrix} \begin{pmatrix} \hat{Y} & 0 & 0 & 0 \\ 0 & -\hat{Y} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} [\cdot]^\top \succ 0,$$

$$\begin{pmatrix} \hat{Y} & I \\ I & \hat{X} \end{pmatrix} \succ 0.$$

The LMIs have the following interpretation with respect to the transformed plant \hat{P} : the first LMI is the primal dissipativity constraint, the second is the dual dissipativity constraint, and the third is the coupling condition [33], [34].

D. Convex synthesis over causal Zames–Falb multipliers

The matrix inequalities in the previous section are linear in the unknowns when the multiplier ψ is fixed, but become nonlinear when the multiplier coefficients are included as variables. While there is no known method to efficiently search over general multipliers, a particularly clever approach was discovered in [9] for the case of causal Zames–Falb multipliers. The convexification procedure hinges on the fact that SISO LTI systems commute, with their state-space realizations related through a particular state transformation. The convex algorithm synthesis based on this insight adapted to our particular setting as follows.

Given filter parameters $\lambda \in \Lambda_\ell^\rho$, the state transformation relating the commuting LTI system is

$$T = \begin{bmatrix} (L-\mu)^{-1} I_\ell & 0 \\ (L-\mu)^{-1} N & \sum_{j=0}^{\ell-1} \lambda_j \rho^j A_p^{-j} \end{bmatrix} = \begin{bmatrix} T_f \\ T \end{bmatrix},$$

where the matrix $N \in \mathbb{R}^{n_p \times \ell}$ is the solution to the Sylvester equation

$$A_p N - \rho N A_f + \mu B_p C_f = 0. \quad (13)$$

With the annihilator matrix $V^\top = [\rho^{-1} \mu B_p^\top \quad L - \mu]_\perp$,

consider the following matrix inequalities:

$$[\cdot]^\top \begin{pmatrix} \hat{X} & 0 & 0 & 0 \\ 0 & -\hat{X} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{bmatrix} \left(\begin{array}{c|c} \hat{A} & \hat{B}_w \\ \hline \hat{C}_z & \hat{D}_{zw} \end{array} \right) \\ \left(\begin{array}{c|c} 0 & 1 \end{array} \right) \end{bmatrix} \hat{U} \prec 0, \quad (14a)$$

$$\begin{bmatrix} V \left(\begin{array}{c|c} -I & A_p \\ \hline 0 & 0 \end{array} \middle| \begin{array}{c} 0 \\ \hat{B}_w \end{array} \right) \\ \left(\begin{array}{c|c} 0 & \hat{D}_{zw} \end{array} \right) \end{bmatrix} \begin{pmatrix} \tilde{Y} & 0 & 0 & 0 \\ 0 & -\rho^{-2} \tilde{Y} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} [\cdot]^\top \succ 0, \quad (14b)$$

$$\begin{pmatrix} \tilde{Y} & \tilde{T} \\ \tilde{T}^\top & \hat{X} \end{pmatrix} \succ 0. \quad (14c)$$

We can now state the main convex synthesis result from [9].

Theorem 2: Fix the rate $\rho \in (0, 1)$, number of multipliers $\ell \in \mathbb{N}$, and model $S \in \mathbb{R}^{p \times p}$. The following are equivalent.

- 1) There exists an algorithm (4) such that the analysis LMI (11) is feasible for some \mathcal{X} and $\lambda \in \Lambda_\ell^\rho$.
- 2) There exist matrices $N \in \mathbb{R}^{n_p \times \ell}$, $\hat{X} \in \mathbb{S}^{n_p+\ell}$, and $\tilde{Y} \in \mathbb{S}^{n_p}$ as well as parameters $\lambda \in \Lambda_\ell^\rho$ that satisfy the Sylvester equation (13) and LMIs (14). \square

The (equivalent) statements in Theorem 2 imply that there exists a controller K for which the corresponding algorithm (7) exponentially tracks the optimizer of (1) for all functions f and trajectories θ that satisfy Assumptions 1 to 3. Moreover, (13) and (14) are *convex* constraints in all of the unknowns, and hence provide an efficient procedure to verify the existence of an algorithm that exponentially tracks the optimizer with a given rate. The minimal rate can then be found by bisection over ρ . The corresponding controller can be reconstructed from the multiplier by finding a solution to the analysis condition (11), thereby solving Problem 1.

V. NUMERICAL INVESTIGATION OF THE METHODOLOGY

We now use the convex synthesis procedure to simulate the synthesized algorithm on a particular time-varying optimization problem and study how the rate depends on parameters.

A. Simulation

We first illustrate our results by applying our designed algorithm to a particular time-varying optimization problem. We consider the instance of (1) with objective

$$f(z, \theta_k) = \frac{1}{2} (z - e_1^\top \theta_k)^2 + a \log(1 + e^{bz}), \quad (15)$$

where $f : \mathbb{R} \times \mathbb{R}^2 \rightarrow \mathbb{R}$ and $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^p$. This models a logistic regression problem with a time-varying regularization term. Intuitively, an optimizer is a point that tracks the time-varying signal $e_1^\top \theta_k$ while avoiding large values, which are penalized by the logistic term. This objective satisfies Assumption 1 with parameters $\mu = 1$ and $L = 1 + \frac{ab^2}{4}$. For our experiments, we used $a = 1$ and $b = 6$.

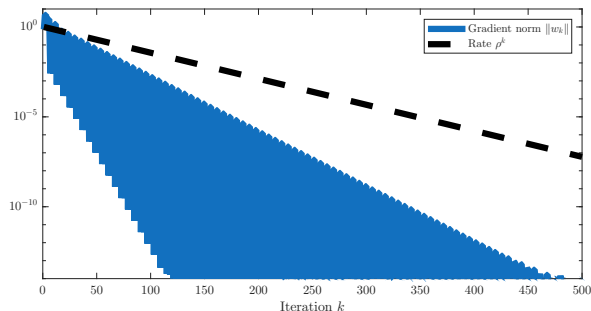


Fig. 5. Simulation of the synthesized controller on a time-varying regularized logistic regression problem; see Section V-A.

We generated the parameter vector θ_k with frequencies $\theta = 0$ and $\theta = \pi/3$. The harmonics are those in Example 1, and the transfer function in (7) is $H(z) = z^5/(z^6 - 1)$. The transfer function of the synthesized controller (after reduction) is $K(z) = -0.18/z^5$. Interestingly, the controller canceled all five zeros at $z = 0$ of $H(z)$, resulting in the closed-loop transfer function $G(z) = -0.18/(z^6 - 1)$, which exponentially tracks the time-varying optimizer of (15) with rate $\rho = 0.9672$. In Figure 5, we plot the gradient norm $\|w_k\|$ of the algorithm trajectory along with the theoretical rate ρ^k , illustrating that the synthesized algorithm attains the expected rate.

B. Convergence rates

The convergence rate of the algorithm depends on the objective parameters μ and L , the exosystem S , and the length ℓ of the multiplier used to characterize the gradient. To better understand how the rate varies with these parameters, we plot ρ as a function of the exosystem frequency θ in Figure 6 with $\mu = 1$, $L = 10$, and $\ell = 1$. The controller is then synthesized using the harmonics Ω_λ , where $\lambda = (e^{j\theta}, e^{-j\theta})$ if $0 < \theta < \pi$, $\lambda = 1$ if $\theta = 0$, and $\lambda = -1$ if $\theta = \pi$. Also shown is the rate of the triple momentum method [8], which achieves the optimal rate $\rho_{\text{TM}} = 1 - \sqrt{\mu/L}$ in the time-invariant case when $\theta = 0$.

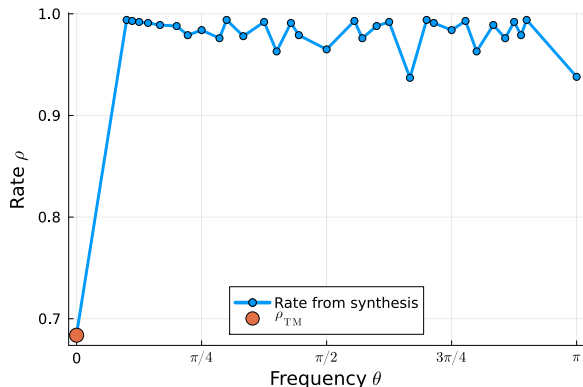


Fig. 6. Rate ρ of the synthesized algorithm as a function of exosystem frequency θ with $\mu = 1$, $L = 10$, and $\ell = 1$.

We now draw several observations from Figure 6:

- The attainable rate varies with the frequency θ of the time variation. More precisely, the rate is not monotonic in θ , meaning that in some cases there exists an algorithm that converges faster for problems with higher frequencies than for lower frequencies.
- Developing further on the point above, the fact that the attainable rate is not constant in θ reveals a key difference from [35], where, for quadratic objectives, lower bounds on the convergence rate do *not* depend on the frequency of the temporal variation, but only on the order of the exosystem model.
- The rate is significantly slower than the time-invariant counterpart (where $\rho = \rho_{\text{TM}} \approx 0.6838$), even for slow time variations (small $\theta > 0$).

VI. CONCLUSIONS

We showed how internal models can be systematically embedded into the convex synthesis framework from robust control to design first-order optimization methods for smooth, strongly convex objectives that vary in time. The proposed methodology is constructive and directly implementable whenever a model of the temporal variability is available, thereby enabling principled algorithm design with guaranteed tracking and robustness properties. Extensions include nonlinear exosystem models and adaptive algorithms that estimate the model online.

APPENDIX I

PROOF OF THEOREM 1

The proof of Theorem 1 relies on two lemmas.

Lemma 1 (Internal model principle): Fix the objective parameters μ and L , exosystem model S , algorithm matrices (A, B, C) , and rate $\rho \in (0, 1)$. Suppose the feedback interconnection of algorithm (4) with uncertainty Δ (see Figure 3 (left)) is robustly locally exponentially stable with rate ρ for all $\Delta \in \mathbf{\Delta}$. Then, the following statements are equivalent:

- Algorithm (4) locally asymptotically tracks the optimizer of (1) with rate ρ for all functions f and exosystem trajectories θ that satisfy Assumptions 1 to 3.
- There exists an analytic mapping $\sigma : \mathbb{R}^p \rightarrow \mathbb{R}^n$ satisfying the internal model kinematic and output equations:

$$\sigma(S\theta) = A\sigma(\theta) \quad \text{and} \quad \nabla_z f(C\sigma(\theta), \theta) = 0.$$

Proof: The result follows from [6, Theorem 2]. \blacksquare

We now explore some consequences of condition (ii). Since σ is real analytic, it admits a (locally convergent) multivariate power series expansion of the form

$$\sigma(\theta) = \sum_{\alpha \in \mathbb{N}^p} c_\alpha \theta^\alpha, \quad \theta^\alpha := \theta_1^{\alpha_1} \cdots \theta_p^{\alpha_p}, \quad (16)$$

where the coefficients $c_\alpha \in \mathbb{R}^n$ are uniquely determined. The monomials $\{\theta^\alpha\}_{\alpha \in \mathbb{N}^p}$ form a linearly independent basis for analytic functions, so this representation is unique.

If S is diagonalizable with eigenvalues $\lambda = (\lambda_1, \dots, \lambda_p)$, then in diagonalized coordinates one has $(S\theta)^\alpha = \lambda^\alpha \theta^\alpha$, where $\lambda^\alpha := \lambda_1^{\alpha_1} \dots \lambda_p^{\alpha_p}$. Substituting the expansion into the kinematic relation $\sigma(S\theta) = A\sigma(\theta)$ yields

$$\sum_{\alpha \in \mathbb{N}^p} \lambda^\alpha c_\alpha \theta^\alpha = \sum_{\alpha \in \mathbb{N}^p} A c_\alpha \theta^\alpha.$$

By linear independence of the monomials, the coefficients of each θ^α must match, which gives

$$(A - \lambda^\alpha I) c_\alpha = 0 \quad \text{for all } \alpha \in \mathbb{N}^p.$$

In addition, by the implicit function theorem, there exists an optimizer $z_*(\theta)$ such that $\nabla_z f(z_*(\theta), \theta) \equiv 0$. Since the optimizer is unique, the output condition requires

$$C \sum_{\alpha} c_\alpha y^\alpha = C \tilde{\sigma}(y) = z_*(y) = \sum_{\alpha} d_\alpha y^\alpha,$$

where $d_\alpha \in \mathbb{R}$ are the coefficients in the series expansion of $z_*(y)$. Again matching coefficients gives $C c_\alpha = d_\alpha$. To summarize, a mapping σ with series expansion (16) satisfies the internal model conditions if and only if $(A - \lambda^\alpha I) c_\alpha = 0$ and $C c_\alpha = d_\alpha$ for each multi-index $\alpha \in \mathbb{N}^p$. This insight leads to the following result on the algorithm structure.

Lemma 2 (Algorithm structure): Fix the objective parameters μ and L , exosystem model S , and algorithm matrices (A, B, C) . Suppose S is diagonalizable with eigenvalues $\lambda = (\lambda_1, \dots, \lambda_p) \in \mathbb{C}^p$. Then, the following are equivalent:

(i) There exists an analytic mapping $\sigma : \mathbb{R}^p \rightarrow \mathbb{R}^n$ satisfying the internal model kinematic and output equations:

$$\sigma(S\theta) = A\sigma(\theta) \quad \text{and} \quad \nabla_z f(C\sigma(\theta), \theta) = 0.$$

(ii) For each harmonic $\omega \in \Omega_\lambda$, A has an eigenvalue at ω and the following non-resonance condition holds:

$$\begin{bmatrix} A - \omega I & B \\ C & 0 \end{bmatrix} \text{ full rank.} \quad \square$$

Proof: (ii) \Rightarrow (i). From (ii), there exists an eigenvector $v_\alpha \in \mathbb{C}^n$ of A with eigenvalue $\lambda^\alpha \in \mathbb{C}$ for each $\alpha \in \mathbb{N}^p$. Moreover, non-resonance implies that the product

$$\begin{bmatrix} A - \lambda^\alpha I & B \\ C & 0 \end{bmatrix} \begin{bmatrix} v_\alpha \\ 0 \end{bmatrix}$$

is nonzero, so $C v_\alpha \neq 0$. From our previous discussion, the mapping σ of the form (16) with $c_\alpha = d_\alpha (C v_\alpha)^{-1} v_\alpha$ satisfies the internal model conditions, so (i) holds.

(i) \Rightarrow (ii). Suppose (i) holds, and consider the series expansion of σ in (16). Suppose that, for some α , the optimizer contains mode y^α , meaning that $d_\alpha \neq 0$. The conditions on c_α imply that it is an eigenvector of A with eigenvalue $\omega = \lambda^\alpha$. Suppose the non-resonance condition fails with this α . Then there exists a nonzero pair (p, q) such that

$$\omega p^\top = p^\top A + q^\top C \quad \text{and} \quad 0 = p^\top B.$$

Multiplying the kinematic internal model condition on the left by p^\top and using this first equation yields

$$p^\top \sigma(S\theta) = \omega p^\top \sigma(\theta) - q^\top z^*(\theta),$$

where $z^*(\theta) = C\sigma(\theta)$ is the optimizer. Now consider a trajectory θ_k of the exosystem and define $y_k := p^\top \sigma(\theta_k)$. This signal evolves according to the linear difference equation

$$y_{k+1} - \omega y_k = -q^\top z^*(\theta_k),$$

which has natural frequency ω . We can choose f such that its optimal trajectory is such that the forcing term $-q^\top z^*(\theta_k)$ is nonzero at this frequency, which then resonates with the system to grow unbounded. But σ is continuous from (i) and the exosystem trajectory is bounded from Assumption 2, so y_k must be bounded — a contradiction. Therefore, the non-resonance condition is satisfied, so (ii) holds. \blacksquare

Item (ii) of Lemma 2 enforces that the algorithm transfer function $G(z)$ must have a pole at each harmonic $\omega \in \Omega_\lambda$. Therefore, any such transfer function must factor as (7) for some proper rational transfer function $K(z)$. With this structure, Problem 1 becomes that of designing the controller $K(z)$ such that (4) is robustly exponentially stable for all time-invariant objective functions that satisfy Assumption 1, from which the claim in Theorem 1 follows.

REFERENCES

- [1] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM J Optim*, vol. 26, no. 1, pp. 57–95, 2016.
- [2] B. Hu and L. Lessard, "Dissipativity theory for Nesterov's accelerated method," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1549–1557.
- [3] C. Scherer and C. Ebenbauer, "Convex synthesis of accelerated gradient algorithms," *SIAM Journal on Control and Optimization*, vol. 59, no. 6, pp. 4615–4645, 2021.
- [4] C. W. Scherer, C. Ebenbauer, and T. Holicki, "Optimization algorithm synthesis based on integral quadratic constraints: A tutorial," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 2995–3002.
- [5] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends in Optimization*, vol. 2, no. 3–4, pp. 157–325, 2016.
- [6] G. Bianchin and B. Van Scoy, "The discrete-time internal model principle of time-varying optimization: Limitations and algorithm design," in *Proc CDC*, Dec. 2025.
- [7] —, "The internal model principle of time-varying optimization," *IEEE Trans Automatic Ctrl*, Aug. 2026, arXiv:2407.08037.
- [8] B. Van Scoy, R. A. Freeman, and K. M. Lynch, "The fastest known globally convergent first-order method for minimizing strongly convex functions," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 49–54, 2017.
- [9] C. W. Scherer, C. Ebenbauer, and T. Holicki, "Optimization algorithm synthesis based on integral quadratic constraints: A tutorial," in *Proc CDC*, 2023, pp. 2995–3002.
- [10] M. Muehlebach and M. Jordan, "A dynamical systems perspective on nesterov acceleration," in *International Conference on Machine Learning*, 2019, pp. 4656–4662.
- [11] C. W. Scherer, "Dissipativity and integral quadratic constraints: Tailored computational robustness tests for complex interconnections," *IEEE Control Systems Magazine*, vol. 42, no. 3, pp. 115–139, 2022.
- [12] L. Lessard, "The analysis of optimization algorithms: A dissipativity approach," *IEEE Control Systems Magazine*, vol. 42, no. 3, pp. 58–72, 2022.
- [13] L. Lessard and P. Seiler, "Direct synthesis of iterative algorithms with bounds on achievable worst-case convergence rate," in *2020 American Control Conference*, 2020, pp. 119–125.
- [14] S. Michalowsky, C. Scherer, and C. Ebenbauer, "Robust and structure exploiting optimisation algorithms: an integral quadratic constraint approach," *International Journal of Control*, vol. 94, pp. 1–24, 2021.

- [15] D. Gramlich, C. Ebenbauer, and C. W. Scherer, "Synthesis of accelerated gradient algorithms for optimization and saddle point problems using Lyapunov functions and LMIs," *Systems & Control Letters*, vol. 165, p. 105271, 2022.
- [16] T. Holicki and C. W. Scherer, "Algorithm design and extremum control: Convex synthesis due to plant multiplier commutation," in *60th IEEE Conference on Decision and Control*, 2021, pp. 3249–3252.
- [17] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *International conference on machine learning*, 2003, pp. 928–936.
- [18] E. Hazan, A. Agarwal, and S. Kale, "Logarithmic regret algorithms for online convex optimization," *Machine Learning*, vol. 69, no. 2, pp. 169–192, 2007.
- [19] M. Colombino, E. Dall'Anese, and A. Bernstein, "Online optimization as a feedback controller: Stability and tracking," *IEEE Trans Control Netw Syst*, vol. 7, no. 1, pp. 422–432, 2020.
- [20] A. Bernstein, E. Dall'Anese, and A. Simonetto, "Online primal-dual methods with measurement feedback for time-varying convex optimization," *IEEE Trans Signal Processing*, vol. 67, no. 8, pp. 1978–1991, 2019.
- [21] A. Simonetto, A. Mokhtari, A. Koppel, G. Leus, and A. Ribeiro, "A class of prediction-correction methods for time-varying convex optimization," *IEEE Trans Signal Processing*, vol. 64, no. 17, pp. 4576–4591, 2016.
- [22] M. Fazlyab, S. Paternain, V. M. Preciado, and A. Ribeiro, "Prediction-correction interior-point method for time-varying convex optimization," *IEEE Trans Automatic Ctrl*, vol. 63, no. 7, pp. 1973–1986, 2017.
- [23] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, "Time-varying convex optimization: Time-structured algorithms and applications," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2032–2048, 2020.
- [24] N. Bastianello, R. Carli, and S. Zampieri, "Internal model-based online optimization," *IEEE Trans Automatic Ctrl*, vol. 69, no. 1, pp. 689–696, 2024.
- [25] G. Bianchin, J. Cortés, J. I. Poveda, and E. Dall'Anese, "Time-varying optimization of LTI systems via projected primal-dual gradient flows," *IEEE Trans Control Netw Syst*, vol. 9, no. 1, pp. 474–486, Mar. 2022.
- [26] B. A. Francis and W. M. Wonham, "The internal model principle of control theory," *Automatica*, vol. 12, no. 5, pp. 457–465, 1976.
- [27] J. Hepburn and W. Wonham, "Error feedback and internal models on differentiable manifolds," *IEEE Trans Automatic Ctrl*, vol. 29, no. 5, pp. 397–403, 1984.
- [28] D. Astolfi, L. Praly, and L. Marconi, "Harmonic internal models for structurally robust periodic output regulation," *Systems and Control Letters*, vol. 161, pp. 105–154, 2022.
- [29] F. Delli Priscoli, "Robust tracking for a class of nonlinear plants, achieved via a linear controller," in *Proceedings of 32nd IEEE Conference on Decision and Control*, 1993, pp. 3550–3555 vol.4.
- [30] C. Scherer and C. Ebenbauer, "Convex synthesis of accelerated gradient algorithms," *SIAM Journal on Control and Optimization*, vol. 59, no. 6, pp. 4615–4645, Jan. 2021.
- [31] C. W. Scherer and C. Ebenbauer, "A tutorial on convex design of optimization algorithms by integral quadratic constraints," *Annual Review of Control, Robotics, and Autonomous Systems*, 2025.
- [32] G. Zames and P. Falb, "Stability conditions for systems with monotone and slope-restricted nonlinearities," *SIAM Journal on Control*, vol. 6, no. 1, pp. 89–108, 1968.
- [33] P. Gahinet and P. Apkarian, "A linear matrix inequality approach to h_∞ control," *International Journal of Robust and Nonlinear Control*, vol. 4, no. 4, pp. 421–448, 1994.
- [34] J. C. Willems, "Dissipative dynamical systems—Part I: General theory; Part II: Linear systems with quadratic supply rates," *Archive for Rational Mechanics and Analysis*, vol. 45, no. 5, pp. 321–393, 1972.
- [35] B. Van Scoy and G. Bianchin, "Temporal variabilities limit convergence rates in gradient-based online optimization," in *American Control Conference*, 2026.