2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

A Framework for Time-Varying Optimization via Derivative Estimation

Matteo Marchi[1], Jonathan Bunton[1], João Pedro Silvestre[2], and Paulo Tabuada[1]

Abstract— **Optimization algorithms have a rich and fundamental relationship with ordinary differential equations given by its continuous-time limit. When the cost function varies with time – typically in response to a dynamically changing environment – online optimization becomes a continuous-time trajectory tracking problem. To accommodate these time variations, one typically requires some inherent knowledge about their nature such as a time derivative.**

In this paper, we propose a novel construction and analysis of a continuous-time derivative estimation scheme based on "dirty-derivatives", and show how it naturally interfaces with continuous-time optimization algorithms using the language of ISS (Input-to-State Stability). More generally, we show how a simple Lyapunov redesign technique leads to provable suboptimality guarantees when composing this estimator with any well-behaved optimization algorithm for time-varying costs.

I. Introduction

Optimization problems form the basis of a vast array of engineering problems [21]. While optimization with respect to a static cost function is a classical field of study, more and more applications are being deployed in dynamic environments, leading to real-time unpredictable changes in the optimization cost function [3], and any small improvement in this setting is extremely beneficial to a variety of tasks.

Optimization algorithms are commonly analyzed in the discrete-time setting. However, numerous classical algorithms such as gradient descent, Nesterov acceleration [14], and their variants can be viewed as discretization schemes of continuous-time ODEs (Ordinary Differential Equations) [2], [18]. When modeled as such, properties of these algorithms (like convergence and stability) can be analyzed via the many tools of control and nonlinear systems theory, such as Lyapunov stability [10]. This view point also impacts the study of time-varying or online optimization. Under some convexity assumptions, time-varying optimization can be analyzed as a *trajectory tracking* problem [19] by determining how closely the resulting solution $x(t)$ tracks the ideal time-varying minimizer trajectory $x^*(t) = \operatorname{argmin}_x f(x,t)$, a widely studied problem in continuous-time control theory.

In the time-varying setting, the literature generally takes one of two approaches. The first is to pick an algorithm developed for the time-invariant case and proving that it is robust to some level of time-variation of the cost function. In [16] the author derives tracking error bounds for the classical gradient descent method, and [24] gives regret bounds in the Online Convex Optimization (OCO) setting under slow variations. The second approach, instead, tries to redesign or adapt an optimization method by directly incorporating knowledge of the time variation. In continuous-time, multiple works add a time-varying correction term to the continuous-time analogue of Newton's method, as under ideal conditions that fully compensates for the time variation [5], [17], [19], with some works also analyzing time-varying constraints [4].

As one might expect, incorporating knowledge of the time variations in the cost can often bring massive performance/stability benefits. However, this typically requires having direct knowledge of the time-variation in the form of a time derivative, such as $\frac{\partial}{\partial t}\nabla_x f(x,t)$. This may not constitute a problem in some applications, such as a control task where we want to track a known reference trajectory, but many others do not allow for this kind of knowledge ahead of time. In [20], this time derivative is approximated via finite differences, but their analysis is tailored to their specific discrete-time algorithm and hard to generalize. Another option, pursued in this work, is to study a general derivative estimation technique, and under what conditions the interconnection of such an estimator with an optimization algorithm results in a well-behaved system.

Derivative estimation has a rich history, from numerous variants of finite difference [6], [11], [15] to interpolation-based [12] schemes. Notably, many observers and state estimation methods from control theory can be adapted to perform derivative estimation (related to the notion of flatness for dynamical systems [7]), of which high-gain observers [1], [9] are a notable example. The so-called dirty-derivative originates from PID control [8], [13] as a common approximation to implement a "derivative" term in the control signal, but has received little interest from a theoretical point of view. This filter is the basis of the extended derivative estimator we introduce in this work. The ability of this construction to interface with the framework we develop is crucial for our analysis.

Our contribution is summarized as follows: 1) We introduce a framework to adapt continuous-time static optimization algorithms to include time-varying knowledge, and establish when this is possible through the language of ISS (Input-to-State Stability) theory [22]. 2) We provide a novel construction of a general derivative estimator based on dirty-derivatives, and we derive explicit IOS (Input-to-Output Stability) [23] estimation error bounds. 3) We establish that

* This work was funded in part by the Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196.

[1]Matteo Marchi, Jonathan Bunton, and Paulo Tabuada are with the Electrical and Computer Engineering Department, University of California at Los Angeles, Los Angeles, CA 90095 USA (e-mail: {matmarchi, j.bunton, tabuada}@ucla.edu).

[2]João Pedro Silvestre is in between IST and UCLA joao.pedro.silvestre@tecnico.ulisboa.pt

the dynamical system resulting from the interconnection of such an adapted optimization algorithm and our derivative estimator results, also, is an IOS system. This translates to robust behavior and bounds on the time-varying minimizer tracking error $\|x(t) - x^*(t)\|$.

II. Problem Statement

We consider a convex optimization task where the cost function varies with time. More explicitly, consider a feasible set $\mathcal{X} \subseteq \mathbb{R}^n$, a set of parameter vectors $\Theta \subseteq \mathbb{R}^p$ and a time domain $\mathbb{T} \subseteq \mathbb{R}_{\geq 0}$, alongside a cost function $f : \mathcal{X} \times \Theta \to \mathbb{R}$. Assuming that the parameters change with time according to some unknown function $\theta : \mathbb{T} \to \Theta$, we seek to design an Ordinary Differential Equation (ODE):

$$\dot{x} = g(x) + u, \tag{1}$$

with $x \in \mathcal{X}$, $g : \mathcal{X} \to \mathbb{R}^n$, and $u : \mathbb{T} \to \mathbb{R}^n$ such that its solution, $x(t)$, minimizes the time-varying function f for all $t \in \mathbb{T}$. Because the time-varying law of the parameters is unknown, g and u cannot be designed using knowledge of differential properties of θ, such as its time derivative $\dot{\theta}(t)$, but we may use its current value $\theta(t)$.

We make the following common assumptions:

Assumption 1: The function f is smooth and strongly convex for every fixed $\overline{\theta} \in \Theta$. I.e., there exists $\mu \in \mathbb{R}_{>0}$ such that:

$$f(x, \overline{\theta}) \geq f(x', \overline{\theta}) + \nabla_x f(x', \overline{\theta})^\top (x - x') + \frac{\mu}{2}\|x - x'\|_2^2,$$

for all $x, x' \in \mathcal{X}$, and $\theta \in \Theta$.

Under these conditions, there exists a (unique) minimizing trajectory described as $x^*(\theta(t)) = \operatorname{argmin}_{x \in \mathcal{X}} f(x, \theta(t))$, and we then seek to minimize a *tracking error* with respect to this trajectory. I.e., we design g and u in (1) such that $x - x^*$ converges to a neighborhood of the origin, the size of which may depend on the magnitude of the time-variations in f.

III. Online Optimization

Classically, continuous-time algorithms in the form (1) arise as the continuous limit of iterative algorithms, with the prime example being the *gradient flow* $\dot{x} = \nabla_x f(x)$. Multiple works have shown the effectiveness of gradient flows, even in online optimization tasks, when the time variations are slow [16].

A related family of methods directly incorporates knowledge of the time-variations in the cost function. A typical candidate for this approach is the continuous limit of Newton's method:

$$\dot{x} = -\left(\nabla_{xx} f(x, \theta)\right)^{-1} \nabla_x f(x, \theta) + u. \tag{2}$$

Here $\nabla_{xx} f(x, \theta)$ denotes the Hessian matrix associated with a time-varying $f(x, \theta)$. When augmenting (2) with time variation knowledge, one often considers the correction:

$$u(t) = -\left(\nabla_{xx} f(x, \theta)\right)^{-1} \nabla_{x\theta} f(x, \theta)\dot{\theta}, \tag{3}$$

where $\nabla_{x\theta} f(x, \theta)$ is the partial derivative of the gradient of f with respect to θ. For a detailed analysis of (2)-(3), see [4].

Traditionally, the performance of (1) is analyzed purely through the lens of stability, but here we consider the slightly more general view of *input-to-state stability* (ISS) [22]. For a fixed θ, consider the corresponding minimizer $x^*(\theta)$. Then, we say (1) is ISS with respect to u if:

$$\|x(t) - x^*(\theta)\| \leq \beta(\|x(t_0) - x^*(\theta)\|, t - t_0) + \kappa(\|u(t)\|_\infty), \tag{4}$$

for all $t_0 \leq t \in \mathbb{T}$, where κ is a class $\mathcal{K}$ function, β is a class $\mathcal{KL}$ function, $\|\cdot\|$ denotes the usual vector 2-norm, and $\|u(t)\|_\infty$ is the essential supremum of u over $[t_0, t]$.[1]

ISS has an equivalent Lyapunov characterization. Equation (4) is equivalent to the existence of a radially unbounded function $V : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}_{\geq 0}$ with $V(x^*(\theta), \theta) = 0$ and $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathcal{K}$ that satisfy:

$$\begin{gathered} \alpha_1(\|x - x^*\|) \leq V(x, \theta) \leq \alpha_2(\|x - x^*\|), \\ \dot{V}(x, \theta) = \nabla_x V(x, \theta)\dot{x} \leq -\alpha_3(V(x, \theta)) + \alpha_4(\|u(t)\|). \end{gathered} \tag{5}$$

Note that when $u = 0$, the system is (uniformly) globally asymptotically stable.

We observe that, under some mild conditions, any ODE (1) that satisfies this ISS property for the family of *time-invariant* optimization problems, can be adapted into a robust ODE for the *time-varying* optimization problem, and formalize this assertion.

Lemma 1: Consider system (1) and suppose g is such that for any fixed value of $\theta \in \Theta$ and its corresponding minimizer $x^*(\theta)$, system (1) is ISS with respect to u. Let $V(x, \theta)$ be a corresponding family of Lyapunov functions differentiable with respect to θ with bounded $\|\nabla_\theta V(x, \theta)\| < \infty$. Then, if $\theta(t)$ is a smooth function of time, any continuous $u : \mathbb{T} \to \mathbb{R}^n$ satisfying:

$$\nabla_x V(x, \theta) u + \nabla_\theta V(x, \theta)(\dot{\theta} + d) \leq 0, \tag{6}$$

renders (1) ISS with respect to the signal $d : \mathbb{T} \to \mathbb{R}^p$ when considering the time-varying minimizer $x^*(\theta(t))$. I.e., the following inequality holds for some $\beta \in \mathcal{KL}$ and $\kappa \in \mathcal{K}$:

$$\left\|x(t) - x^*(\theta(t))\right\| \leq \beta\left(\left\|x(0) - x^*(\theta(0))\right\|, t\right) + \kappa(\|d\|_\infty). \tag{7}$$

In later sections the signal d will be used to model the estimation error on $\dot{\theta}$.

Proof: We follow the same basic steps as in Lyapunov redesign for ISS [22]. Taking the time derivative of $V(x, \theta)$ we have:

$$\begin{aligned} \dot{V}(x, \theta) &= \nabla_x V(x, \theta)(g(x) + u) + \nabla_\theta V(x, \theta)\dot{\theta} \\ &= \nabla_x V(x, \theta) g(x) + \nabla_x V(x, \theta) u + \nabla_\theta V(x, \theta)\dot{\theta} \\ &\leq -\alpha_3(V(x, \theta)) + \nabla_x V(x, \theta) u + \nabla_\theta V(x, \theta)\dot{\theta} \\ &\leq -\alpha_3(V(x, \theta)) - \nabla_\theta V(x, \theta) d \\ &\leq -\alpha_3(V(x, \theta)) + \alpha_d(\|d\|), \end{aligned}$$

[1] A class $\mathcal{K}$ function $\kappa : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is strictly increasing with $\kappa(0) = 0$. A class $\mathcal{KL}$ function $\beta : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a continuous function such that $\beta(\cdot, s) \in \mathcal{K}$ and $\beta(r, \cdot)$ is strictly decreasing with $\lim_{s \to \infty} \beta(r, s) = 0$.

where we have used (5) in the first inequality and defined $\alpha_d(\|d\|) = \sup_{x,\theta}(\|\nabla_\theta V(x,\theta)\|)\|d\|$. Then, the existence of β and κ satisfying (7) immediately follows by equivalence with the Lyapunov characterization of ISS. ■

This approach applies generically to any continuous-time algorithm that is asymptotically stable for the entire class of *static* problems encountered, such as strongly convex costs. In this way, Lemma 1 provides a framework for adapting a sufficiently well-behaved continuous-time algorithm to the time-varying scenario.

Remark 1: A choice of u satisfying (6) always exists as long as $\nabla_x V(x,\theta) \neq 0$ whenever $\nabla_\theta V(x,\theta) \neq 0$. As a special case, this is always satisfied for the choice[2] $V(x,\theta) = \frac{1}{2}\|\nabla_x f(x,\theta)\|^2$. Under Assumption 1 on f, this is a Lyapunov function for (2) and the correction (3) always satisfies condition (6) with $d = 0$.

Remark 2: If condition (6) is not always satisfied, we can still obtain a weaker version of Lemma 1 as long as there is a finite threshold $\overline{V} \in \mathbb{R}_{\geq 0}$ such that $V(x,\theta) > \overline{V}$ implies $\nabla_x V(x,\theta) \neq 0$. In that case the result reduces to ISpS (input-to-state practical stability).

IV. Derivative Estimation

In Lemma 1, we showed how online optimization methods can be designed by exploiting knowledge of the time variations in the function. Unfortunately, the class of control inputs in (6) is characterized by explicit knowledge of $\dot{\theta}$, which is generally unavailable in the online scenario. To remedy this issue, we consider an extension of the classical continuous-time differentiation scheme, sometimes called *dirty-derivative* [13], and provide a proof of convergence.

The dirty-derivative operator approximates the derivative of a signal $w : \mathbb{T} \to \mathbb{R}$ and is described by the transfer function:

$$\frac{\sigma s}{s+\sigma}, \tag{8}$$

where $\sigma \in \mathbb{R}_{>0}$ determines how fast the filter defined by the transfer function tracks the derivative of w.

We generalize (8) to estimate derivatives up to some order k, additionally resulting in a flexible choice of error bounds. The proposed estimator follows a recursive structure, where $\widehat{W}_i(s)$ is the Laplace transform of the i-th estimated derivative for $i \in \{1, 2, \ldots, k\}$:

$$\begin{aligned} \widehat{W}_i(s) &= \frac{\sigma^i s^i}{(s+\sigma)^i} W(s) + \frac{(s+\sigma)^i - \sigma^i}{s(s+\sigma)^i} \widehat{W}_{i+1}(s) \\ \widehat{W}_k(s) &= \frac{\sigma^k s^k}{(s+\sigma)^k} W(s). \end{aligned} \tag{9}$$

A block diagram of (9) for $k = 3$ is shown in Fig. 1. Each derivative estimated by (9) converges exponentially to a ball around the true derivative whose size is a function of the magnitude of the $(k+1)$-th derivative of w.

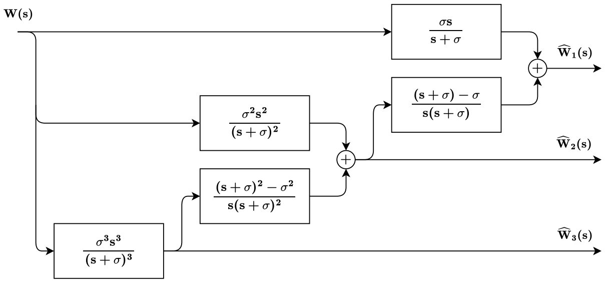
$\mathbf{W(s)}$
$\frac{\sigma \mathbf{s}}{\mathbf{s}+\sigma}$
$\widehat{\mathbf{W}}_1\mathbf{(s)}$
$+$
$\frac{(\mathbf{s}+\sigma)-\sigma}{\mathbf{s}(\mathbf{s}+\sigma)}$
$\frac{\sigma^2\mathbf{s}^2}{(\mathbf{s}+\sigma)^2}$
$+$
$\widehat{\mathbf{W}}_2\mathbf{(s)}$
$\frac{(\mathbf{s}+\sigma)^2-\sigma^2}{\mathbf{s}(\mathbf{s}+\sigma)^2}$
$\frac{\sigma^3\mathbf{s}^3}{(\mathbf{s}+\sigma)^3}$
$\widehat{\mathbf{W}}_3\mathbf{(s)}$

Fig. 1. System (9) estimating the first three derivatives of a signal $U(s)$.

We are now in the position to prove our previous claim for the dirty-derivative-based system (9). Note that we use $w^{(i)}$ to denote the i-th derivative of w.

Theorem 1: Let $w : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a $k+1$ times differentiable input signal and $\widehat{w}_i : \mathbb{R}_{\geq 0} \to \mathbb{R}$ for $i = \{1, 2, \ldots, k\}$ be the i-th estimate produced by (9) when fed w. Then, for all $t \geq 0$:

$$\left\|\widehat{w}_i(t) - w^{(i)}(t)\right\| \leq \beta_{i,\sigma} e^{-\gamma_i \sigma t} + \frac{\alpha_i}{\sigma^{k+1-i}} \left\|w^{(k+1)}(t)\right\|_\infty, \tag{10}$$

for some $\beta_{i,\sigma} \in \mathbb{R}_{>0}$ dependent on initial conditions, and $\gamma_i, \alpha_i \in \mathbb{R}_{>0}$. In particular:

$$\lim_{t\to\infty} \|\widehat{w}_i(t) - w^{(i)}(t)\| \leq \frac{\alpha_i}{\sigma^{k+1-i}} \|w^{(k+1)}(t)\|_\infty. \tag{11}$$

Proof: Let us define the estimation error of the i-th derivative of w as $e_i = \hat{w}_i - w^{(i)}$. Then, for $i = k$ the Laplace transform of the error is:

$$\begin{aligned} E_k = \widehat{W}_k - s^k W &= \frac{\sigma^k s^k}{(s+\sigma)^k} W - s^k W \\ &= \frac{\sigma^k - (s+\sigma)^k}{(s+\sigma^k)} s^k W = -\frac{(s+\sigma)^k - \sigma^k}{s(s+\sigma^k)} s^{k+1} W \\ &= -F_k(s) s^{k+1} W, \end{aligned} \tag{12}$$

where we define $F_i(s)$ for $i \in \{1, 2, \ldots, k\}$ as:

$$F_i(s) = \frac{(s+\sigma)^i - \sigma^i}{s(s+\sigma)^i}. \tag{13}$$

For $i < k$ the error is:

$$\begin{aligned} E_i &= \widehat{W}_i - s^i W \\ &= \frac{\sigma^i s^i}{(s+\sigma)^i} W + \frac{(s+\sigma)^i - \sigma^i}{s(s+\sigma)^i} \widehat{W}_{i+1} - s^i W \\ &= \frac{\sigma^i - (s+\sigma)^i}{(s+\sigma^i)} s^i W + \frac{(s+\sigma)^i - \sigma^i}{s(s+\sigma)^i} \widehat{W}_{i+1} \\ &= \frac{(s+\sigma)^i - \sigma^i}{s(s+\sigma)^i} \left(\widehat{W}_{i+1} - s^{i+1} W\right) \\ &= \frac{(s+\sigma)^i - \sigma^i}{s(s+\sigma)^i} E_{i+1} \\ &= F_i(s) E_{i+1}. \end{aligned} \tag{14}$$

Because the transfer function $F_i(s)$ corresponds to the state-space system in Lemma 2 (see the appendix[3]), we can apply

[2] A simple computation shows that $\nabla_x V(x,\theta) = \nabla_{xx} f(x,\theta) \nabla_x f(x,\theta)$ and $\nabla_\theta V(x,\theta) = \nabla_{x\theta} f(x,\theta) \nabla_x f(x,\theta)$. Therefore, $\nabla_\theta V \neq 0$ implies $\nabla_x f \neq 0$ which in turns implies $\nabla_x V \neq 0$ because of strong convexity of f.

[3] Inspecting the system defined in Lemma 2, this can be easily verified by computing $C_n(sI - A_n)^{-1} B_n$. Because $A_{n,\sigma}$ is in controllable canonical form, $\det{(sI - A_n)^{-1}} = (s+\sigma)^n$, and due to C_n and B_n, we only need to compute the top left cofactor of $sI - A_n$, which equals $s^{-1}\left((s+\sigma)^n - \sigma^n\right)$.

its result recursively (denoting the internal state of F_i as x_{e_i}):

$$
\begin{aligned}
\|e_i(t)\| &\le \|x_{e_i}(0)\| b_{i,\sigma} e^{-d_i \sigma t} + \frac{a_i}{\sigma}\|e_{i+1}(t)\| \\
&\le \|x_{e_i}(0)\| b_{i,\sigma} e^{-d_i \sigma t} \\
&\quad + \frac{a_i}{\sigma}\left(\|x_{e_{i+1}}(0)\| b_{i+1,\sigma} e^{-d_{i+1}\sigma t} + \frac{a_{i+1}}{\sigma}\|e_{i+2}(t)\|\right) \\
&\;\vdots \\
&\le \sum_{j=i}^{k}\left(\left(\prod_{\ell=i}^{j-1}\frac{a_\ell}{\sigma}\right)\|x_{e_j}(0)\| b_{j,\sigma} e^{-d_j \sigma t}\right) \\
&\quad + \left(\prod_{j=i}^{k} a_j\right)\frac{\|w^{(k+1)}(t)\|_\infty}{\sigma^{k+1-i}}.
\end{aligned}
$$

Because the first term is a sum of exponentially decaying terms with a common σ in the exponent, there exist $\beta_{i,\sigma}, \gamma_i, \alpha_i \in \mathbb{R}_{>0}$ such that:

$$
\begin{aligned}
&\sum_{j=i}^{k}\left(\left(\prod_{\ell=i}^{j-1}\frac{a_\ell}{\sigma}\right)\|x_{e_j}(0)\| b_{j,\sigma} e^{-d_j \sigma t}\right) \le \beta_{i,\sigma} e^{-\gamma_i \sigma t}, \\
&\left(\prod_{j=i}^{k} a_j\right)\frac{\|w^{(k+1)}(t)\|_\infty}{\sigma^{k+1-i}} \le \alpha_i \frac{\|w^{(k+1)}(t)\|_\infty}{\sigma^{k+1-i}}.
\end{aligned}
\tag{15}
$$

■

***Remark** 3:* Although this result is for a scalar signal, it immediately generalizes to estimating k-derivatives of an m-dimensional signal $w : \mathbb{T} \to \mathbb{R}^m$ by running m copies of the estimator in parallel; one for each scalar component of w.

***Remark** 4:* The asymptotic error bound in Theorem 1 approaches zero for $\sigma \to \infty$, but with a rate that scales differently with the order of derivative $i \le k$. In particular, higher-order derivatives are more difficult to estimate, and the size of the ball to which we are guaranteed to converge scales inversely with powers of the derivative order.

***Remark** 5:* In the special case where the signal $w : \mathbb{T} \to \mathbb{R}$ is described by a k-degree polynomial, the $(k+1)$-th derivative vanishes, and $\left\|w^{(k+1)}(t)\right\|_\infty = 0$. Theorem 1 then guarantees that the estimates converge exponentially to the exact values of the derivatives of w for any choice of σ.

V. Framework for Time-varying Optimization

Both the online optimization algorithms designed via Lemma 1 and the dirty derivative scheme in Theorem 1 come with ISS-style convergence guarantees (specifically, Theorem 1 gives an IOS (input-to-output stability) result [23]). Because of these results, we can cascade these systems and preserve their convergence properties. In particular, Lemma 1 shows that a well-behaved (in the ISS sense) time-invariant optimization algorithm can be adapted into a time-varying one assuming knowledge of $\dot{\theta}$. We now present our main contribution, and show that the estimates produced by our dirty-derivative scheme (9) can be used in place of $\dot{\theta}$ while still preserving input-to-output stability of the interconnected system. We formalize this notion in the following theorem.

***Theorem** 2:* Consider a system (1) satisfying the assumptions of Lemma 1, and let $\widehat{\theta}_1 : \mathbb{T} \to \Theta$ be the estimate of $\dot{\theta}$ provided by a dirty-derivative estimator (9) of order k when fed θ as input. Then, if u is selected such that:

$$
\nabla_x V(x,\theta) u + \nabla_\theta V(x,\theta)\widehat{\theta}_1 \le 0, \tag{16}
$$

the combination of system (1) and the estimator is IOS (input-to-output stable) with respect to the signal $\sigma^{-k}\theta^{(k+1)}(t)$ when considering the time-varying minimizer $x^*(\theta(t))$. I.e., with outputs $\widetilde{x}(t) = x(t) - x^*(\theta(t))$ and $e_1 = \widehat{\theta}_1 - \dot{\theta}$ there exist $\beta'_\sigma \in \mathcal{KL}$ and $\kappa' \in \mathcal{K}$ such that:

$$
\begin{aligned}
&\left\|\left(\widetilde{x}(t), e_1(t)\right)\right\| \le \\
&\quad \beta'_\sigma\left(\left\|\left(\widetilde{x}(0), x_e(0)\right)\right\|, t\right) + \kappa'\left(\sigma^{-k}\left\|\theta^{(k+1)}\right\|_\infty\right),
\end{aligned}
\tag{17}
$$

where x_e denotes the full internal state of the dirty-derivative error system.

Proof: We first observe that by (16):

$$
\begin{aligned}
0 &\ge \nabla_x V(x,\theta) u + \nabla_\theta V(x,\theta)\widehat{\theta}_1 \\
&= \nabla_x V(x,\theta) u + \nabla_\theta V(x,\theta)(\dot{\theta} + (\widehat{\theta}_1 - \dot{\theta})) \\
&= \nabla_x V(x,\theta) u + \nabla_\theta V(x,\theta)(\dot{\theta} + e_1).
\end{aligned}
\tag{18}
$$

Then, by applying Lemma 1 with $d = e_1$ we can write:

$$
\left\|x(t) - x^*(\theta(t))\right\| \le \beta\left(\left\|x(0) - x^*(\theta(0))\right\|, t\right) + \kappa(\|e_1\|_\infty),
$$

and:

$$
\begin{aligned}
\left\|\left(\widetilde{x}(t), e_1(t)\right)\right\| &\le \|\widetilde{x}(t)\| + \|e_1(t)\| \\
&= \left\|x(t) - x^*(\theta(t))\right\| + \|e_1(t)\| \\
&\le \beta\left(\left\|x(0) - x^*(\theta(0))\right\|, t\right) + \kappa(\|e_1\|_\infty) \\
&\quad + \beta_{1,\sigma} e^{-\gamma_1 \sigma t} + \frac{\alpha_1}{\sigma^k}\left\|\theta^{(k+1)}(t)\right\|_\infty.
\end{aligned}
$$

By the nonlinear superposition principle [22, Section 3.1], only the essential supremum norm for $t \to \infty$ matters, so we can substitute the essential supremum norm $\|e_1(t)\|_\infty$ by a limsup $\overline{\lim}_{t\to\infty}\|e_1(t)\| \le \frac{\alpha_1}{\sigma^k}\left\|\theta^{(k+1)}(t)\right\|_\infty$:

$$
\begin{aligned}
\left\|\left(\widetilde{x}(t), e_1(t)\right)\right\| &\le \left[\beta\left(\left\|\widetilde{x}(0)\right\|, t\right) + \beta_{1,\sigma} e^{-\gamma_1 \sigma t}\right] \\
&\quad + \left[\kappa\left(\frac{\alpha_1}{\sigma^k}\left\|\theta^{(k+1)}(t)\right\|_\infty\right) + \frac{\alpha_1}{\sigma^k}\left\|\theta^{(k+1)}(t)\right\|_\infty\right].
\end{aligned}
$$

By composition of class $\mathcal{K}$ functions, and because $\beta_{1,\sigma} e^{-\gamma_1 \sigma t}$ is monotonically decaying with time with $\beta_{1,\sigma}$ dependent on the initial conditions of the internal error system's state, there exist $\beta'_\sigma \in \mathcal{KL}$ and $\kappa' \in \mathcal{K}$ such that (17) holds.

■

Theorem 2, together with Lemma 1 and our dirty-derivative estimator, provide a general framework to adapt a continuous time-invariant optimization algorithm to an online time-varying algorithm. Notably, explicit time derivative knowledge of θ is not necessary, since we can approximate it while retaining the stability properties of the ideal algorithm, and the performance of the optimization algorithm can be improved arbitrarily by increasing the gain σ. However, in practice the presence of measurement noise on θ, or a discrete implementation of the continuous time algorithm will limit how high σ can be set. The effect of noise on our proposed construction is explored empirically in the following section.

VI. Experiments

In this section we demonstrate in simulation the effectiveness of our proposed derivative estimation scheme (9), and its application to a time-varying optimization task where we use it to estimate the time-varying correction in Lemma 1.

A. Derivative Tracking

We first show the system (9) tracking the first derivative of a sinusoidal signal $w(t) = \sin(5t-2)$ for $\sigma \in \{5, 20\}$ (see Fig. 2). As expected, the tracking is significantly better for the bigger value of σ. In a practical scenario, we may only have access to a corrupted version of w, perturbed by some noise $n(t)$, i.e., $\widetilde{w}(t) = w(t) + n(t)$, and we want our estimation to reject this noise effectively. Noise rejection is practically relevant even when the signal is fully measurable, as a typical time-discretization of (9) is equivalent to the continuous system working with bounded noise.

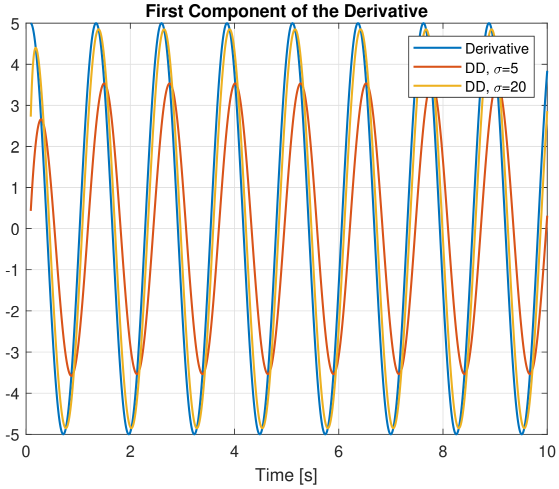

Fig. 2. Plot of the first derivative of the signal $w(t) = \sin(5t-2)$ and its estimates for $\sigma \in \{5, 20\}$.

In Fig. 3, we show first-derivative tracking of the same signal $w(t) = \sin(5t-2)$ for $\sigma \in \{5, 20\}$ corrupted by zero mean Gaussian noise with $\operatorname{var}(n) = 0.01$.

Even with high σ, the estimator closely tracks the true derivative while rejecting most of the high frequency noise.

B. Time-Varying Optimization with Derivative Estimator

We now show a simple application of (9) in a time-varying optimization context. We pick a time-varying loss function:

$$f(x,\theta) = \frac{1}{2}\|x - \theta(t)\|^2,$$

$$\theta(t) = \begin{bmatrix}\cos(5t-2) & \sin(5t-2) & \cos^2(5t-2)\end{bmatrix}^T,$$

that we minimize using the continuous-time generalization of Newton's method (2) with the time-varying correction (3).

In the simulation, we consider an *estimated* correction $\widehat{u}(t) = -\left(\nabla_{xx} f(x,\theta)\right)^{-1} \nabla_{x\theta} f(x,\theta)\widehat{\theta}_1$ where we use our dirty derivative construction (9) to estimate $\dot{\theta}(t)$ from just the available signal $\theta(t)$. In Fig. 4 we compare our achieved loss over time with the loss achieved using the ideal correction (3).

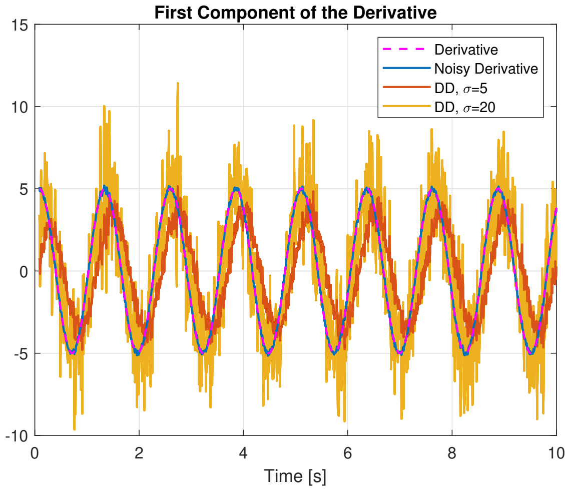

Fig. 3. Plot of the first derivative of the signal $w(t) = \sin(5t-2)$ and its estimates for $\sigma \in \{5, 20\}$ when w is corrupted by Gaussian noise.

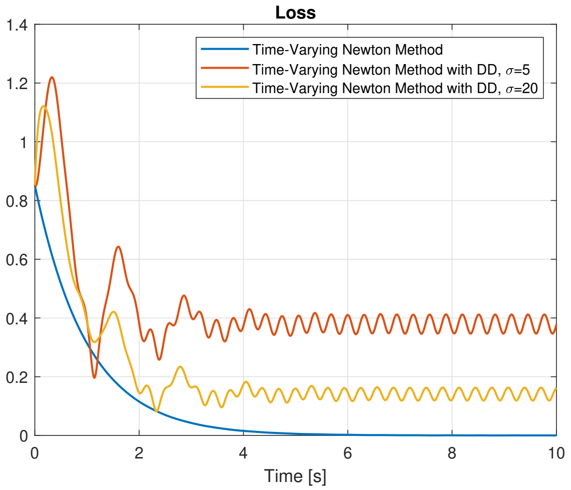

Fig. 4. Loss (over time) of the online Newton's method with ground truth (blue) and estimated (red, yellow) time derivative information.

Fig. 4 shows that our loss trajectory approximates the ideal one, settling closer to zero loss with higher values of σ. Practically, however, higher values of σ tend to amplify the effects of noisy measurements, preventing us from scaling σ arbitrarily high. We repeat the experiment with the derivative estimator being fed a signal $\theta(t)$ with each component corrupted by Gaussian noise with zero mean and 0.01 variance. In Fig. 5 the experiment results show again good robustness to the added noise.

VII. Conclusions

We have presented a framework to adapt a time-invariant continuous-time optimization algorithm to incorporate a correction term compensating for the time-variation, and showed that popular methods, such as (2), are suitable for this. We then provided a novel general purpose construction of a derivative estimator with explicit error bounds, and showed that using these estimates in place of true time derivative

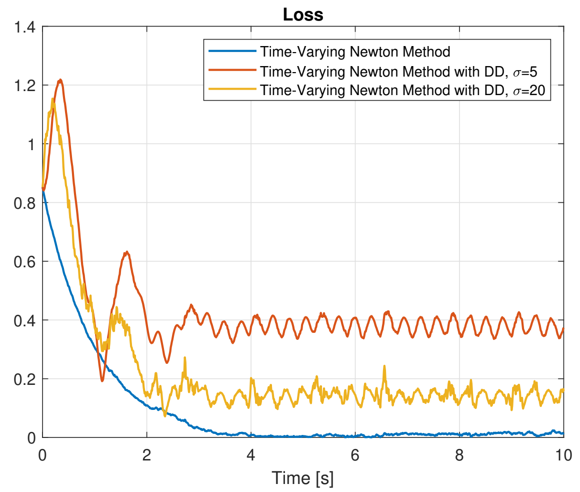

Fig. 5. Loss (over time) of the online Newton's method with noise-corrupted ground truth (blue) and estimated (red, yellow) time derivative information.

knowledge in the adapted algorithm results in an IOS system. This provides robustness and convergence results for the tracking error with respect to the true time-varying minimizer of the cost function. Finally, we verified our theory via a set of proof-of-concept simulation experiments.

In future studies we intend to extend the current framework and results to not only the time-varying parameter problem $\theta(t)$ studied here, but to the full closed-loop interaction with a solution dependent parameter vector $\theta(x(t))$.

Appendix

Lemma 2: Consider the LTI SISO dynamical system:

$$\begin{cases} \dot{x}(t) &= A_{n,\sigma}x(t) + B_n u(t) \\ y(t) &= C_n x(t), \end{cases} \tag{19}$$

where $x : \mathbb{T} \to \mathbb{R}^n$, $u, y : \mathbb{T} \to \mathbb{R}$, $\sigma \in \mathbb{R}_{>0}$, and the matrices $A_{n,\sigma} \in \mathbb{R}^{n\times n}$, $B_n \in \mathbb{R}^{n\times 1}$, and $C_n \in \mathbb{R}^{1\times n}$ are:

$$A_{n,\sigma} = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ -\binom{n}{0}\sigma^n & -\binom{n}{1}\sigma^{n-1} & \dots & -\binom{n}{n-1}\sigma \end{bmatrix},$$

$$B_n = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T, \qquad C_n = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}.$$

Then, there exist $b_{n,\sigma}, d_n, a_n \in \mathbb{R}_{>0}$ such that for all $t \in \mathbb{T}$, the output y of (19) satisfies:

$$\|y(t)\| \leq \|x(0)\| b_{n,\sigma} e^{-d_n \sigma t} + \frac{a_n}{\sigma}\|u(t)\|_\infty. \tag{20}$$

For length considerations, we omit the tedious proof of this technical lemma. The full proof can be found in the extended version of the paper (`https://arxiv.org/abs/2403.19088`).

References

[1] Daniele Astolfi, Lorenzo Marconi, Laurent Praly, and Andrew R Teel. Low-power peaking-free high-gain observers. *Automatica*, 98:169–179, 2018.

[2] Pedro Cisneros-Velarde and Francesco Bullo. A contraction theory approach to optimization algorithms from acceleration flows. In *International Conference on Artificial Intelligence and Statistics*, pages 1321–1335. PMLR, 2022.

[3] Emiliano Dall'Anese, Andrea Simonetto, Stephen Becker, and Liam Madden. Optimization and learning with information streams: Time-varying algorithms and applications. *IEEE Signal Processing Magazine*, 37(3):71–83, 2020.

[4] Mahyar Fazlyab, Santiago Paternain, Victor M Preciado, and Alejandro Ribeiro. Prediction-correction interior-point method for time-varying convex optimization. *IEEE Transactions on Automatic Control*, 63(7):1973–1986, 2017.

[5] Ping Gong, Fei Chen, and Weiyao Lan. Time-varying convex optimization for double-integrator dynamics over a directed network. In *2016 35th Chinese Control Conference (CCC)*, pages 7341–7346. IEEE, 2016.

[6] Hassan Zohair Hassan, AA Mohamad, and GE Atteia. An algorithm for the finite difference approximation of derivatives with arbitrary degree and order of accuracy. *Journal of Computational and Applied Mathematics*, 236(10):2622–2631, 2012.

[7] Alberto Isidori, CH Moog, and A De Luca. A sufficient condition for full linearization via dynamic state feedback. In *1986 25th IEEE Conference on Decision and Control*, pages 203–208. IEEE, 1986.

[8] Michael A Johnson and Mohammad H Moradi. *PID control*. Springer, 2005.

[9] Hassan K Khalil and Laurent Praly. High-gain observers in nonlinear feedback control. *International Journal of Robust and Nonlinear Control*, 24(6):993–1015, 2014.

[10] H.K. Khalil. *Nonlinear Systems*. Macmillan Publishing Company, 1992.

[11] Ishtiaq Rasool Khan and Ryoji Ohba. Closed-form expressions for the finite difference approximations of first and higher derivatives based on taylor series. *Journal of Computational and Applied Mathematics*, 107(2):179–193, 1999.

[12] Sunder Ram Krishnan and Chandra Sekhar Seelamantula. On the selection of optimum savitzky-golay filters. *IEEE transactions on signal processing*, 61(2):380–391, 2012.

[13] Antonio Loría. Observers are unnecessary for output-feedback control of Lagrangian systems. *IEEE Transactions on Automatic Control*, 61(4):905–920, 2015.

[14] Michael Muehlebach and Michael Jordan. A dynamical systems perspective on nesterov acceleration. In *International Conference on Machine Learning*, pages 4656–4662. PMLR, 2019.

[15] Kailash C Patidar. On the use of nonstandard finite difference methods. *Journal of Difference Equations and Applications*, 11(8):735–758, 2005.

[16] A Yu Popkov. Gradient methods for nonstationary unconstrained optimization problems. *Automation and Remote Control*, 66:883–891, 2005.

[17] Salar Rahili and Wei Ren. Distributed continuous-time convex optimization with time-varying cost functions. *IEEE Transactions on Automatic Control*, 62(4):1590–1605, 2017.

[18] Orlando Romero, Mouhacine Benosman, and George J Pappas. ODE discretization schemes as optimization algorithms. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 6318–6325. IEEE, 2022.

[19] Andrea Simonetto, Emiliano Dall'Anese, Santiago Paternain, Geert Leus, and Georgios B Giannakis. Time-varying convex optimization: Time-structured algorithms and applications. *Proceedings of the IEEE*, 108(11):2032–2048, 2020.

[20] Andrea Simonetto, Aryan Mokhtari, Alec Koppel, Geert Leus, and Alejandro Ribeiro. A class of prediction-correction methods for time-varying convex optimization. *IEEE Transactions on Signal Processing*, 64(17):4576–4591, 2016.

[21] Ramteen Sioshansi, Antonio J Conejo, et al. Optimization in engineering. *Cham: Springer International Publishing*, 120, 2017.

[22] Eduardo D. Sontag. *Input to State Stability: Basic Concepts and Results*, pages 163–220. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[23] Eduardo D Sontag and Yuan Wang. Notions of input to output stability. *Systems & Control Letters*, 38(4-5):235–248, 1999.

[24] Tianbao Yang, Lijun Zhang, Rong Jin, and Jinfeng Yi. Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient. In *International Conference on Machine Learning*, pages 449–457. PMLR, 2016.