Online Optimization with Feedback
Andrey Bernstein, Emiliano Dall’Anese, Andrea Simonetto*

Abstract—This paper addresses the design and analysis of feedback-based online algorithms to control systems or networked systems based on performance objectives and engineering constraints that may evolve over time. The emerging time-varying convex optimization formalism is leveraged to model optimal operational trajectories of the systems, as well as explicit local and network-level operational constraints. Departing from existing batch and feed-forward optimization approaches, the design of the algorithms capitalizes on an online implementation of primal-dual projected-gradient methods; the gradient steps are, however, suitably modified to accommodate actionable feedback from the system – hence, the term online optimization with feedback. By virtue of this approach, the resultant algorithms can cope with model mismatches in the algebraic representation of the system states and outputs, it avoids pervasive measurements of exogenous inputs, and it naturally lends itself to a distributed implementation.

Under suitable assumptions, analytical convergence claims are established in terms of dynamic regret. Furthermore, when the synthesis of the feedback-based online algorithm is based on a regularized Lagrangian function, Q-linear convergence to solutions of the time-varying optimization problem is shown.

I. INTRODUCTION

This paper focuses on time-varying optimization problems [1] associated with systems or networked systems, for the purpose of modeling and controlling their operation based on performance objectives and engineering constraints that may evolve over time [2]–[5]. The term “networked systems” here refers to a collection of systems coupled through intrinsic physical and behavioral interdependencies, and logically connected by an information infrastructure that supports given network-level control and optimization tasks. Examples include power grids, transportation networks, water systems, and robotic networks just to mention a few [6].

Suppose that physical and/or behavioral interdependencies among systems in the network are modeled as

$$y(t) = M(x(t); t)$$

(1)

where $$x(t) \in \mathbb{R}^n$$ is a vector collecting given controllable inputs of the systems, $$y(t) \in \mathbb{R}^m$$ represents observables or outputs of the network (quantities that pertain to both edges and nodes), and $$M(\cdot; t) : \mathbb{R}^n \rightarrow \mathbb{R}^m$$ is a time-varying map defined over the domain of $$x(t)$$. For example, when a linear network model is utilized, (1) boils down to

$$y(t) = Cx(t) + Dw(t)$$

(2)

where $$C \in \mathbb{R}^{m \times n}$$ and $$D \in \mathbb{R}^{m \times w}$$ are given model parameters, and $$w(t) \in \mathbb{R}^w$$ is a vector of time-varying exogenous inputs (or, simply, uncontrollable quantities in the network).

Consider associating with the networked systems a time-varying optimization of the form

$$\min_{x \in X(t)} f(x, y(x); t)$$

(3)

where $$t \in \mathbb{R}_+$$ is the temporal index; $$X(t)$$ is a convex set; $$f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+ \rightarrow \mathbb{R}$$ is a convex function at each time $$t$$; and, the notation $$y(x; t)$$ is utilized to stress that the observables $$y(t)$$ depend on the vector variable $$x$$. The function $$f$$ is time-varying, in the sense that it can capture performance objectives that evolve over time. Accordingly, denoting as $$x^*(t)$$ an optimal solution of (3) at time $$t$$, the optimization model (3) leads to a continuous-time optimal trajectory. Given (1) and (3), the problem addressed in this paper pertains to the development of algorithms that enable tracking of the optimal trajectory $$\{x^*(t)\}_{t \in \mathbb{R}_+}$$.

For an isolated system or when the map (1) does not depend on time-varying exogenous inputs that are geographically and logically dispersed in the network, problem (3) might be solved in a centralized setting based on a continuous time platform (see e.g., [3], [4], [7], [8]; however, this paper focuses on the case where (i) the algorithmic framework is distributed across systems, (ii) measurements and communication of exogenous inputs introduce non-negligible delays, and (iii) the update of the input $$x(t)$$ leads to control actions that are implemented on digital control units.

Let $$s > 0$$ denote a given sampling time and consider discretizing the solution trajectory of (3) as $$\{x^*(t_k)\}_{k \in \mathbb{N}}$$, where $$t_k := k s$$. For perfect tracking, (3) can be re-interpreted as a sequence of time-invariant problems that must be solved to convergence (i.e., batch solution) at each time $$t_k$$. However, a batch solution of (3) might not be achievable within a interval that is consistent with the variability of $$f(x; t)$$ and the map $$M(\cdot; t)$$ due to underlying communication and computational complexity requirements; for example, since iterative methods require multiple computation and communication rounds, the problem inputs $$f(\cdot; t)$$ and $$M(\cdot; t)$$ (and therefore the solution) might have already changed by the time the iterative method converged. Consider then the following online first-order algorithm, tailored to the model (2) and to the case where the cost is $$f(y(x; t); t)$$ for exposition simplicity:

$$x(t_{k+1}) = \text{Proj}_{X(t_k)} \left\{ x(t_k) - \alpha C^T \nabla_x f(Cx(t_k) + Dw(t_k); t_k) \right\}$$

(4)

1 Notation: Upper-case (lower-case) boldface letters will be used for matrices (column vectors), and $$(\cdot)^T$$ denotes transposition. For a given $$N \times 1$$ vector $$x \in \mathbb{R}^N$$, $$\|x\|_2 := \sqrt{x^T x}$$. Given a matrix $$X \in \mathbb{R}^{N \times M}$$, $$[X]_{m,n}$$ denotes its $$(m,n)$$-th entry and $$[X]_m$$ denotes the $$m$$-th induced matrix norm. For a function $$f : \mathbb{R}^N \rightarrow \mathbb{R}$$, $$\nabla_x f(x)$$ returns the gradient vector of $$f(x)$$ with respect to $$x \in \mathbb{R}^N$$. Finally, $$\text{proj}_{X}(x)$$ denotes a closest point to $$x$$ in $$X$$, namely $$\text{proj}_{X}(x) \in \arg\min_{x \in X} \|x - y\|_2$$.

* Alphabetical order, authors contributed equally to the paper.

A. Bernstein and E. Dall’Anese are with the National Renewable Energy Laboratory, Golden, CO, USA. A. Simonetto is with the IBM Research Dublin, Dublin, Ireland. Emails: andrey.bernstein@nrel.gov, emiliano.dallanese@nrel.gov, andrea.simonetto@ibm.com.
where \( \text{Proj}_X(z) := \arg \min_{x \in X} \|z - x\|_2 \) denotes projection onto a convex set and \( \alpha > 0 \) is the step size. It is clear that \( s \), in this case, represents the time required to perform one algorithmic iteration.

Even before elaborating on possible tracking properties of \( \mathcal{A} \), it is important to emphasize that the update \( \mathcal{A} \) represents a feed-forward (i.e., open loop) control method that presumes knowledge of the input-output map \( \mathcal{M} \). In fact, the function \( f(\cdot; t) \) in \( \mathcal{A} \) is evaluated at the current output of the network, based on the postulated model \( y(t_k) = Cx(t_k) + Dw(t_k) \). From a real-time optimization and control perspective, this feature has fundamental drawbacks:

(i) The update \( \mathcal{A} \) requires one to estimate the exogenous inputs \( w(t_k) \) at each time \( t_k \); this is impractical in many existing networked systems, especially when the number of exogenous inputs \( w \) is much larger than \( n \) and \( m \) or when (part of) \( w(t_k) \) might not be even observable.

(ii) The feed-forward strategy \( \mathcal{A} \) is sensitive to model mismatches; errors in the map \( \mathcal{M} \) might drive the network operation to points that might not be implementable.

(iii) The mathematical structure of the map \( \mathcal{M}(x(t); t) \) may prevent a distributed implementation of the update \( \mathcal{A} \).

(iv) The update \( \mathcal{A} \) does not acknowledge that the underlying systems may be governed by local controllers with given state dynamics; in fact, \( \mathcal{A} \) presumes a time-scale separation where the local systems settle to a steady-state in response to a new command \( x(t_k) \) within an interval \( s \).

To address challenges outlined above, the idea suggested in this paper is to suitably modify online optimization methods, such as \( \mathcal{A} \), to accommodate actionable feedback – hence, the term online optimization with feedback. In particular, letting \( \tilde{x}(t_k) \) and \( \tilde{y}(t_k) \) be measurements of the input \( x(t_k) \) and the output \( \tilde{y}(t_k) \), respectively, we consider modifying \( \mathcal{A} \) as

\[
\mathcal{x}(t_k+1) = \text{Proj}_{\mathcal{X}(t_k+1)} \left\{ \tilde{x}(t_k) - \alpha C^T \nabla_y f(\tilde{y}(t_k); t) \right\}
\]

where the measurement \( \tilde{y}(t_k) \) replaces the network model and \( \tilde{x}(t_k) \) may replace the current iterate \( x(t_k) \). This simple conceptual modification leads to the following key advantages:

(a.1) Instead of measuring/estimating \( w \) exogenous inputs \( w(t_k) \), \( \mathcal{A} \) relies on \( m \) measurements of the outputs \( \tilde{y}(t_k) \). This is of key importance when \( m \ll w \).

(a.2) The algorithm naturally accounts for the network physics via the measurements \( \tilde{y}(t_k) \), and it does not rely on a synthetic network model.

(a.3) The update \( \mathcal{A} \) may naturally lend itself to a distributed implementation. And,

(a.4) The update \( \mathcal{A} \) accounts for imperfect implementations/commands of the input \( x(t_k) \) at the local systems.

While the simplified setting \( \mathcal{A} \) and \( \mathcal{A} \) was adopted to outline the main ideas, the following sections will present a much broader framework applicable to time-varying constrained convex problems. The design of the algorithms capitalizes on an online implementation of primal-dual projected-gradient methods; however, similar to \( \mathcal{A} \), the gradient steps are suitably modified to accommodate measurements. When the feedback-based primal-dual gradient method is applied to the time-varying Lagrangian, a dynamic regret analysis \( \mathcal{A} \) is provided.

On the other hand, when considering a regularized Lagrangian function \( \mathcal{A} \), \( \mathcal{A} \), performance of the proposed methods is assessed in terms of convergence of the iterates \( \mathcal{x}(t_k) \) within a ball centered around the optimal trajectory \( \{x^*(t_k)\}_{k \in N} \).

This paper generalizes the domain-specific technical findings of \( \mathcal{A} \), \( \mathcal{A} \), \( \mathcal{A} \) in three different ways: (i) it considers generic time-varying constrained convex optimization problems (in contrast, \( \mathcal{A} \), \( \mathcal{A} \) deal with linearly-constrained problems and \( \mathcal{A} \) leverages relaxations via approximate barrier functions); (ii) it addressed the case where feedback is included in both primal and dual gradient steps; (iii) it provides a dynamic regret analysis when a primal-dual gradient method is applied to the time-varying Lagrangian function; and, (iv) it generalizes the Q-linear convergence analysis of \( \mathcal{A} \), \( \mathcal{A} \) when the algorithm is synthesized based on a regularized Lagrangian function \( \mathcal{A} \), \( \mathcal{A} \). As a byproduct, the paper provides contributions over, e.g., \( \mathcal{A} \), \( \mathcal{A} \), where static optimization problems were considered and the earlier work \( \mathcal{A} \) where no analytical convergence results were provided.

In terms of existing literature on regret analysis for online dual and primal-dual methods \( \mathcal{A} \)–\( \mathcal{A} \), the contributions consist in: (i) proving dynamic (as opposed to static) regret bounds; (ii) considering a general class of constrained optimization problems with feedback; (iii) assuming time-varying feasible sets; and (iv) providing a bound on the average constraint violation. The recent works \( \mathcal{A} \), \( \mathcal{A} \) developed frameworks and associated regret bounds that are close in spirit to bounds provided in this paper. The main contribution of this paper compared to these works is to consider the feedback-based optimization and provide regret bounds under slightly weaker (i.e., less restrictive) assumptions; however, this paper does not provide asymptotic zero-regret bounds as this is out of scope of the present paper (cf. Remark \( \mathcal{A} \) in Section \( \mathcal{A} \)).

From an optimization standpoint, the paper extends the results of primal-dual-type methods of e.g., \( \mathcal{A} \), \( \mathcal{A} \), \( \mathcal{A} \) to the case of time-varying problems and when feedback is utilized in the algorithmic steps [cf. \( \mathcal{A} \)]. With respect to the time-varying problem formulations in \( \mathcal{A} \)–\( \mathcal{A} \), the paper provides results in the case of feedback-based methods. It is also worth pointing out that the proposed methodology can be cast within the domain of ε-gradient methods \( \mathcal{A} \)–\( \mathcal{A} \); in this case, the paper extends the analysis of ε-gradient methods to time-varying settings. Lastly, the paper provides an extensions of saddle-point flows \( \mathcal{A} \), \( \mathcal{A} \), \( \mathcal{A} \), \( \mathcal{A} \), \( \mathcal{A} \) to the case of discrete-time steps, time-varying saddle functions, and feedback-based algorithmic steps.

The development of feedback-based online optimization methods has been, so far, driven by power systems application; see, for example, the works on frequency control \( \mathcal{A} \), \( \mathcal{A} \) for transmission systems and for explicit power control in \( \mathcal{A} \), \( \mathcal{A} \). However, the framework is generally applicable to a number of settings where the objective is to drive the operation of physical and logical systems as well as networked systems to optimal operating points in real time. Application domains include, for example, wireless communication systems \( \mathcal{A} \), \( \mathcal{A} \), vehicle control \( \mathcal{A} \), \( \mathcal{A} \), water systems \( \mathcal{A} \), and robotic sensor networks \( \mathcal{A} \).

The remainder of paper is organized as follows. Section \( \mathcal{A} \)
will formulate the time-varying optimization problem and will outline the proposed feedback-based online algorithm. Section III will provide a regret analysis for the algorithm when applied to the Lagrangian function, while Section IV will focus on regularized Lagrangian functions. Section V concludes the paper.

II. FEEDBACK-BASED PRIMAL-DUAL METHOD

Consider a network of $N$ systems, with the associated time-varying optimization problem:

$$(P0)^{(t)} \min_{x \in \mathbb{R}^n} f_0(y(x;t);t) + \sum_{i=1}^{N} f_i(x_i;t)$$

subject to: $x_i \in \mathcal{X}_i(t)$, $i = 1, \ldots, N$

$$g_j(y(x;t);t) \leq 0, j = 1, \ldots, M$$

with $\mathcal{X}_i(t) \subset \mathbb{R}^{n_i}; \sum_{i=1}^{N} n_i = n$; and, where $y(x;t) := Cx + Dw(t) \in \mathbb{R}^m$ is an algebraic representation of some observables in the systems as in (2). Function $f_0(y(x;t);t) : \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}$ is convex and captures costs associated with the outputs $y(x;t)$, while $f_i(x_i;t) : \mathbb{R}^{n_i} \times \mathbb{R}_+ \to \mathbb{R}$ is a convex function that models time-varying costs associated with the $i$-th sub-vector $x_i$. Finally, the convex functions $g_j(y(x;t);t) : \mathbb{R}^m \times \mathbb{R}_+ \to \mathbb{R}$ are utilized to impose time-varying constraints on $y(x;t)$. We assume that $g_j(y(x;t);t)$, for $j = 1, \ldots, M_f$ is nonlinear and convex, whereas $g_j(y(x;t);t)$, for $j = M_f + 1, \ldots, M$, is linear or affine.

As explained in the previous section, consider discretizing the temporal axis as $t_k = ks$, $k \in \mathbb{N}$, where $s > 0$ is a given sampling interval [11, 28]. Accordingly, samples of the continuous-time problem (6) can be expressed as

$$(P0)^{(k)} \min_{x} f_0^{(k)}(y^{(k)}(x)) + \sum_{i=1}^{N} f_i^{(k)}(x_i)$$

subject to: $x_i \in \mathcal{X}_i^{(k)}$, $i = 1, \ldots, N$

$$g_j^{(k)}(y^{(k)}(x)) \leq 0, j = 1, \ldots, M$$

where $\mathcal{X}_i^{(k)} := \mathcal{X}_i(t_k)$, $f_i^{(k)}(x_i) := f_i(x_i;t_k)$, $y^{(k)}(x) = Cx + Dw^{(k)}$, and similar notation is utilized for the remaining sampled quantities. The following assumption is presupposed.

Assumption 1. Slater’s constraint qualification holds at each time instant $k$.

From Assumption 1 it follows that strong duality holds uniformly in time for the convex problems (7). For brevity, define $g^{(k)}(y^{(k)}(x)) := [g_1^{(k)}(y^{(k)}(x)), \ldots, g_M^{(k)}(y^{(k)}(x))]^T$, $f^{(k)}(x) := \sum_{i=1}^{N} f_i^{(k)}(x_i)$ and

$$h^{(k)}(x) := f^{(k)}(x) + f_0^{(k)}(y^{(k)}(x)).$$

Further, let $\lambda \in \mathbb{R}^M$ denote the vector of dual variables associated with (7). Then, the time-varying Lagrangian function is given by:

$$L^{(k)}(x, \lambda) := f^{(k)}(x) + f_0^{(k)}(y^{(k)}(x)) + \lambda^T g^{(k)}(y^{(k)}(x)).$$

Similar to, e.g., [10], consider the following regularized Lagrangian function

$$L_{p,d}^{(k)}(x, \lambda) := L^{(k)}(x, \lambda) + \frac{p}{2} \|x\|^2 - \frac{d}{2} \|\lambda\|^2$$

where $p > 0$ and $d \geq 0$ are given regularization parameters, and consider the following time-varying minimax problem:

$$\max_{\lambda \in \mathcal{X}^{(k)}} \min_{x \in \mathcal{Y}^{(k)}} L_{p,d}^{(k)}(x, \lambda)$$

where $\mathcal{X}^{(k)} := \mathcal{X}_1^{(k)} \times \ldots \times \mathcal{X}_N^{(k)}$ and $\mathcal{Y}^{(k)}$ is a convex and compact set constructed as explained shortly in Section III or as in [10, 24]. Hereafter, $z^{(k)} := \{x^{(k)}, \lambda^{(k)}\}_{k \in \mathbb{N}}$ denote an optimal trajectory of (11).

Based on the time-varying minimax problem (11), the sequential execution of the following steps constitutes the proposed feedback-based online primal-dual gradient algorithm:

$$x^{(k+1)} = \text{Proj}_{\mathcal{X}^{(k)}}\{1 - \alpha\lambda^{(k)}(\alpha \nabla_x f^{(k)}(x^{(k)}))$$

$$+ C^T \nabla_y f_0^{(k)}(\tilde{y}^{(k)}) + \sum_{j=1}^{M} \lambda_j^{(k)} C^T g_j^{(k)}(\tilde{y}^{(k)})\}$$

where $\alpha > 0$ is a constant step size, and $\tilde{y}^{(k)}$ is a measurement of $y^{(k)}(x^{(k)})$ collected at time $t_k$. In the following, convergence results will be provided for the online algorithm (12), depending on the choice of the parameters $p$ and $d$. In particular, the following two cases are in order.

Case 1: $p = 0, d = 0$. Obviously, $L_{0,0}^{(k)}(x, \lambda) = L^{(k)}(x, \lambda)$, and $\{x^{(k)}\}_{k \in \mathbb{N}}$ is a (discretized) optimal solution trajectory of (6). To capture the temporal variability of (11) (and, hence, of (7) as well as its continuous-time counterpart), define the following quantity:

$$\sigma^{(k)} := \|x^{(k+1)} - x^{(k)}\|_2.$$
the Euclidean distance between \( z^{(s,k)} \) and the output of the algorithm \( z^{(k)} := \{ x^{(k)}, \lambda^{(k)} \} \); that is, the following quantity will be bounded:

\[
S^{(k)} := \| z^{(k)} - z^{(s,k)} \|_2. \tag{15}
\]

Section [IV] will show that \( S^{(k)} \) convergences \( Q \)-linearly within a ball centered about the optimal trajectory \( z^{(s,k)} \). To derive bounds on \( S^{(k)} \), the following quantity will be utilized to capture the temporal variability of the optimizer \( z^{(s,k)} \) [cf. (13)]:

\[
\tilde{\sigma}^{(k)} := \| z^{(s,k+1)} - z^{(s,k)} \|_2. \tag{16}
\]

It is worth pointing out that the dynamic regret analysis is applicable also to Case 2; however, the objective of Case 2 is to establish \( Q \)-linear convergence associated with \( S^{(k)} \), which is made possible by the use of a regularized Lagrangian function. These two cases highlight the different convergence results that become available based on the choice of the parameters \( p \) and \( d \).

For exposition simplicity, the paper focuses on the case where only measurements of \( y^{(k)}(x^k) \) are utilized in the steps (12); however, the results can be naturally extended to the case where measurements of \( x^k \) are utilized too.

Pertinent assumptions that are utilized to derive the results explained above are stated next.

**Assumption 2.** The set \( \mathcal{X}^{(k)} \) is convex and compact for all \( k \). Moreover, the sequence \( \{ \mathcal{X}^{(k)} \} \) is uniformly bounded. That is, \( B := \sup_{p \geq 1} \sup_{x \in \mathcal{X}^{(k)}} \| x \|_2 < \infty \). Also, let \( D < \infty \) denote the upper bound on the diameters of \( \{ \mathcal{X}^{(k)} \} \), so that \( \text{diam}(\mathcal{X}^{(k)}) \leq D \) for all \( k \).

**Assumption 3.** The functions \( f_0^{(k)}(y) \) and \( f_1^{(k)}(y) \) are convex and continuously differentiable for all \( k \). The gradient map \( \nabla_x f^{(k)}(x) \) is Lipschitz continuous with constant \( L \geq 0 \) over \( \mathbb{R}^n \) for all \( k \). Furthermore, \( \nabla_y f_0^{(k)}(y) \) is Lipschitz continuous with constant \( L_0 \geq 0 \) over \( \mathbb{R}^m \) for all \( k \).

**Assumption 4.** For each \( j = 1, \ldots, M_j \) and all \( k \), the function \( g_j^{(k)}(y) \) is convex and continuously differentiable. Moreover, it has a Lipschitz continuous gradient with constant \( L_{g,j} > 0 \). Let \( J^{(k)}(y) \) denote the Jacobian (matrix-valued) map of \( g^{(k)}(y) \) with entries

\[
\left( J^{(k)}(y) \right)_{i\ell} := \frac{\partial (g^{(k)}(y))_{i}}{\partial (y)_{\ell}}, \tag{17}
\]

and let \( L_G \geq 0 \) denote the Lipschitz constant of \( J^{(k)}(y) \).

It is worth noticing that, from the continuity of the Jacobian and the compactness of \( \mathcal{X}^{(k)} \), there exists a scalar \( M_g < +\infty \) such that \( \| J^{(k)}(y) \|_2 \leq M_g \) for all \( k \). In fact, one can set:

\[
M_g = \sup_k \max_{y \in \mathcal{X}^{(k)}} \| J^{(k)}(y) \|_2. \tag{18}
\]

Then, using the Mean Value Theorem, one can show that

\[
\| g^{(k)}(y_1) - g^{(k)}(y_2) \|_2 \leq M_g \| y_1 - y_2 \|_2 \tag{19}
\]

for all \( k \in \mathbb{N} \). The parameter \( M_g \) will be utilized in the subsequent sections to establish various convergence results. Since the online algorithm (12) leverages measurements of \( y^{(k)}(x^{(k)}) \) at each time \( k \in \mathbb{N} \), it is appropriate to introduce the following assumption.

**Assumption 5.** There exists a scalar \( e_y < +\infty \) such that the measurement error can be bounded as

\[
\sup_{k \geq 1} \| \hat{y}^{(k)} - y^{(k)}(x^{(k)}) \|_2 \leq e_y. \tag{20}
\]

The bound \( e_y \) can capture finite measurements and quantization errors, model mismatches between the network physics and the algebraic representation (2), and imperfect implementations of the input \( x^{(k)} \) at the local systems/nodes.

With these assumptions in place, a dynamic regret analysis will be presented in the ensuing section. Per-iteration and asymptotic bounds on \( S^{(k)} \) will then be presented in Section IV.

### III. Regret Analysis

Recall that in Case 1 the regularization parameters are \( p = d = 0 \), then, since \( L_{(k)}(x, \lambda) = L^{(k)}(x, \lambda) \), the primal update (12a) can be compactly re-written as

\[
x^{(k+1)} = \text{Proj}_{\mathcal{X}^{(k)}} \{ x^{(k)} - \alpha \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)}) \}, \tag{21}
\]

where

\[
\nabla_x L^{(k)}(x, \lambda) := \nabla f^{(k)}(x) + C^T \nabla f_0^{(k)}(\hat{y}^{(k)})
\]

\[
+ \left( J^{(k)}(\hat{y}^{(k)}) C \right)^T \Lambda \tag{22}
\]

On the other hand, the sets \( D^{(k)} \) in (12b) are conveniently chosen as follows:

\[
D^{(k)} \equiv \Lambda_{\alpha, \kappa} := \left\{ \Lambda \in \mathbb{R}^M_+ : \| \Lambda \|_2 \leq \frac{1}{\alpha^\kappa} \right\} \tag{23}
\]

for some \( \kappa > 0 \). Notice that, for a given step size \( \alpha \), the parameter \( \kappa \) can be chosen so that the set \( D^{(k)} \) contains the convex and compact sets for the dual variables utilized in, e.g., [10], [24]. A similar choice of \( D^{(k)} \) can be utilized in Section IV.

The dynamic regret of the algorithm is analyzed next. To this end, introduce the following notation for brevity:

\[
G := \| C \|_2 M_g, \tag{24}
\]

\[
F := \sup_k \sup_{x \in \mathcal{X}^{(k)}} \| \nabla h^{(k)}(x) \|_2, \tag{25}
\]

\[
g := \sup_k \sup_{x \in \mathcal{X}^{(k)}} \| g^{(k)}(y^{(k)}(x)) \|_2, \tag{26}
\]

\[
L_x := \| C \|_2 \left( L_0 + \frac{L_g}{\alpha^\kappa} \right), \tag{27}
\]

\[
F_x := F + \frac{G}{\alpha^\kappa}. \tag{28}
\]

With this notation in place, the following results for the dynamic regret and constraint violation are presented.

**Theorem 1.** Under Assumptions [1] [2] [3] [4] and [5] for any \( \alpha > 0, \kappa > 0, \) and \( k \in \mathbb{N} \), we have that

\[
R^{(k)} \leq B^{(k)}(\alpha, \kappa) := \frac{B}{\alpha^k} + K_1 \alpha + K_2 \alpha^{1-\kappa} + K_3 \alpha^{1-2\kappa}
\]

\[
+ K_4(\alpha) e_y + K_5(\alpha) e_y^2 + K_6(\alpha) \frac{1}{k} \sum_{\ell=1}^{k} \sigma^{(\ell)} \tag{29}
\]
and, therefore,
\[
\limsup_{k \to \infty} R^{(k)} \leq B^{(\infty)}(\alpha, \kappa) := K_1 \alpha + K_2 \alpha^{1-\kappa} + K_3 \alpha^{1-2\kappa} + K_4(\alpha) e_y + K_5(\alpha) e_y^2 + K_6(\alpha) \limsup_{k \to \infty} \frac{1}{k} \sum_{\ell=1}^{k} \sigma^{(\ell)}
\]
where
\[
K_1 := \frac{1}{2} (F^2 + g^2), \quad K_2 := FG, \quad K_3 := \frac{1}{2} G^2,
\]
\[
K_4(\alpha) := (2B + \alpha F + \alpha^{1-\kappa}G)L_x + \left( \frac{2}{\alpha \kappa} + \alpha g \right) M_y,
\]
\[
K_5(\alpha) := \frac{\alpha}{2} (L_x^2 + M_y^2),
\]
\[
K_6(\alpha) := \frac{1}{\alpha}(D + B).
\]

**Theorem 2.** Under Assumptions [1]  [2]  [3]  [4] and [5] for any \( \alpha > 0, \kappa > 0, \) and \( k \in \mathbb{N}, \) the constraint violation induced by the algorithm [12] can be bounded as:
\[
\frac{1}{k} \sum_{\ell=1}^{k} g^{(\ell)}(y^{(\ell)}(x^{(\ell)})) \leq \alpha^\kappa \left( B^{(k)}(\alpha, \kappa) - R^{(k)} \right)
\]
\[
\leq \alpha^\kappa (B^{(\infty)}(\alpha, \kappa) + 2FB),
\]
and, therefore,
\[
\limsup_{k \to \infty} \frac{1}{k} \sum_{\ell=1}^{k} g^{(\ell)}(y^{(\ell)}(x^{(\ell)})) \leq \alpha^\kappa \left( B^{(\infty)}(\alpha, \kappa) - \liminf_{k \to \infty} R^{(k)} \right)
\]
\[
\leq \alpha^\kappa (B^{(\infty)}(\alpha, \kappa) + 2FB),
\]
where the limsup and inequality are component-wise; and \( B^{(k)}(\alpha, \kappa) \) and \( B^{(\infty)}(\alpha, \kappa) \) are given in (29) and (30), respectively.

The following remarks are in order.

**Remark 1.** The optimal choice of the parameter \( \kappa \) is in general hard to obtain due to the complicated dependency of the terms \( K_1(\alpha) \) and \( K_2(\alpha) \) on \( \kappa. \) Ignoring the terms corresponding to \( e_y, \) the optimal choice is \( \kappa = \frac{1}{\beta} \). Indeed, the dominating term in (30) is \( K_3 \alpha^{1-2\kappa}, \) and the dominating term in (32) is \( 2FB \alpha^\kappa. \) Therefore, asymptotically, for \( \alpha \to 0, \) the optimum is obtained when \( 1 - 2\kappa = \kappa. \)

**Remark 2.** The definition of dynamic regret utilized in [14] is with respect to the optimal sequence \( \{x^{(k)}\}. \) However, the results of Theorems [1] and [2] hold also for any comparator (or reference) sequence \( \{x^{(k)}\}, \) as is for example in [9]. In that case, \( \sigma^{(k)} := \|x^{(k+1)} - \tilde{x}^{(k)}\| \) captures the temporal variability of the comparator sequence.

**Remark 3.** Note that in the error-free case \( (e_y = 0) \) and when the variability of the comparator sequence (cf. Remark [2]) is bounded, namely
\[
\sum_{\ell=1}^{k} \sigma^{(\ell)} \leq B_\sigma, \forall k
\]
for some \( B_\sigma < \infty, \) the obtained results are similar in spirit to the classical dynamic regret bounds (e.g., in [9]). In particular, taking \( \kappa = \frac{1}{\beta} \) as in Remark [1] it follows from (29) and (31) that
\[
R^{(k)} \leq \frac{1}{\alpha k} (B + B_\sigma (D + B)) + K_1 \alpha + K_2 \alpha^{\frac{1}{2}} + K_3 \alpha^{\frac{1}{2}}
\]
\[
\frac{1}{k} \sum_{\ell=1}^{k} g^{(\ell)}(y^{(\ell)}(x^{(\ell)})) \leq \frac{1}{\alpha k} (B + B_\sigma (D + B))
\]
\[
+ K_1 \alpha^{\frac{1}{2}} + K_2 \alpha^{\frac{1}{2}} + K_3 \alpha^{\frac{1}{2}} + 2FB \alpha^{\frac{1}{2}}.
\]
Therefore, using a standard choice of \( \alpha := \frac{1}{k^\beta} \) for some \( 0 < \beta < 1, \) one would obtain
\[
R^{(k)} \leq \frac{1}{k^{1-\beta}} (B + B_\sigma (D + B)) + \frac{K_1}{k^{\frac{1}{2} \beta}} + \frac{K_2}{k^{\frac{1}{4} \beta}} + \frac{K_3}{k^{\frac{1}{4} \beta}},
\]
\[
\frac{1}{k} \sum_{\ell=1}^{k} g^{(\ell)}(y^{(\ell)}(x^{(\ell)})) \leq \frac{1}{k^{1-\beta}} (B + B_\sigma (D + B))
\]
\[
+ \frac{K_1}{k^{\frac{1}{2} \beta}} + \frac{K_2}{k^{\frac{1}{4} \beta}} + \frac{K_3}{k^{\frac{1}{4} \beta}} + 2FB \alpha^{\frac{1}{2}}.
\]
Since the dominating terms are \( \frac{1}{k^{1-\beta}} \) and \( \frac{1}{k^{\frac{1}{2} \beta}}, \) if \( \beta \) satisfies \( 1 - \beta = \frac{1}{2} \beta, \) (namely \( \beta = \frac{2}{3}) \), then one achieves the best convergence rate simultaneously for dynamic regret and constraint violation of \( O(1/k^{\frac{1}{2}}). \) Note that this convergence rate is inferior to the optimal regret bound of \( O(1/\sqrt{k}) \) known in the literature for the standard online convex optimization algorithms; providing algorithms with optimal regret bounds for the problem considered in this paper is a subject of future work.

To prove Theorem [1] and Theorem [2] the following intermediate results are first shown.

**Lemma 1.** For any \( \lambda \in \Lambda_{\alpha, \kappa}, \) the following holds:
\[
\left\| \nabla_x L^{(k)} (x^{(k)}, \lambda) - \nabla_x \hat{L}^{(k)} (x^{(k)}, \lambda) \right\|_2 \leq L_x e_y.
\]
Moreover, \( \| \nabla_x L^{(k)} (x, \lambda) \| \) is uniformly bounded by \( F_x. \)

**Proof.** Note that
\[
\nabla_x L^{(k)} (x, \lambda) := \nabla f^{(k)}(x) + C^T \nabla f_0^{(k)} (y^{(k)}(x)) + \left( J^{(k)} (y^{(k)}(x)) \right) C^T \lambda
\]
By comparing with (22), we have that
\[
\left\| \nabla_x L^{(k)} (x^{(k)}, \lambda) - \nabla_x \hat{L}^{(k)} (x^{(k)}, \lambda) \right\|_2
\]
\[
\leq \|C\|_2 \| \nabla f_0^{(k)} (y^{(k)}(x^{(k)})) - \nabla f_0^{(k)} (\hat{y}^{(k)}) \|_2 + \|\lambda\|_2 \|C\|_2 \|J^{(k)} (y^{(k)}(x^{(k)))) - J^{(k)} (\hat{y}^{(k)})\|_2
\]
\[
\leq \|C\|_2 L_0 \|y^{(k)}(x^{(k)}) - \hat{y}^{(k)}\|_2 + \|\lambda\|_2 \|C\|_2 L_G \|y^{(k)}(x^{(k)}) - \hat{y}^{(k)}\|_2
\]
\[
\leq \left( \|C\|_2 L_0 + \frac{1}{\alpha^\kappa} \|C\|_2 L_G \right) e_y,
\]
where the first inequality holds by the triangle and Cauchy-Schwarz inequalities; the second inequality follows by Assumptions [3] and [4] and the last inequality is due to the fact that \( \lambda \in \Lambda_{\alpha, \kappa} \) and Assumption [5].
Lemma 2. For any $\lambda$, it holds that:
\[
\left\| g^{(k)}(\tilde{x}^{(k)}) - \nabla_{\lambda} L^{(k)}(x^{(k)}, \lambda) \right\|_2 \leq M_g e_y.
\]
Furthermore, $\|\nabla_{\lambda} L^{(k)}(x, \lambda)\|$ is uniformly bounded by $g$.

Proof. The result follows from (19) and Assumption 5.

Lemma 3. For every $k$, the following inequality holds:
\[
L^{(k)}(x^{(k)}, \lambda^{(k)}) - L^{(k)}(x^{(s,k)}, \lambda^{(k)})
\leq \frac{\|x^{(k)} - x^{(s,k)}\|^2 - \|x^{(k+1)} - x^{(s,k+1)}\|^2}{(2\alpha) + \alpha F^2} + \frac{2B + \alpha F_x}{\alpha} \|x^{(s,k)} - x^{(s,k+1)}\|
\]

Furthermore, for any $\lambda \in \Lambda_{\alpha,n}$, it holds that:
\[
\left\| L^{(k+1)}(x^{(k)}, \lambda) - L^{(k)}(x^{(k)}, x^{(s,k+1)}) \right\|^2
\geq \frac{\|\lambda^{(k+1)} - \lambda\|^2 - \|\lambda^{(k)} - \lambda\|^2}{(2\alpha) - \alpha g^2}
\]
\[
- \frac{2}{\alpha^2} \left( g + \frac{M_\gamma e_y}{2} \right)\| \lambda^{(k+1)} - \lambda \|.
\]

Proof. We have that
\[
\|x^{(k+1)} - x^{(s,k+1)}\|^2
= \|x^{(k+1)} - x^{(s,k)} + x^{(s,k)} - x^{(s,k+1)}\|^2
= \|x^{(k+1)} - x^{(s,k)}\|^2 + 2(x^{(k+1)} - x^{(s,k)})^T(x^{(s,k)} - x^{(s,k+1)})
+ \|x^{(s,k)} - x^{(s,k+1)}\|^2
\leq \left\| x^{(k+1)} - x^{(s,k)} - \alpha \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)}) \right\|^2
+ \left[ 2(x^{(k+1)} - x^{(s,k)}) + (x^{(s,k)} - x^{(s,k+1)}) \right]^T(x^{(s,k)} - x^{(s,k+1)})
\leq \|x^{(s,k)} - x^{(s,k+1)}\|^2 + 2(D + B)\|x^{(s,k)} - x^{(s,k+1)}\|,
\]
where the first inequality follows by (21) and the non-expansiveness property of the projection operator; and in the last inequality, we used the Cauchy-Schwarz inequality and the fact that under Assumption 2
\[
\|2(x^{(k+1)} - x^{(s,k)}) + (x^{(s,k)} - x^{(s,k+1)})\|
\leq 2\|x^{(k+1)} - x^{(s,k)}\| + \|x^{(s,k)} - x^{(s,k+1)}\|
\leq 2\text{diam}(X^{(k)}) + \|\lambda^{(k)}\| + \|\lambda^{(k+1)}\|
\leq 2(D + B).
\]

We now expand the first term in (35). It holds that
\[
\|x^{(k)} - x^{(s,k)} - \alpha \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
= \|x^{(k)} - x^{(s,k)} - \alpha \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
+ \alpha \left( \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)}) - \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)}) \right)^2.
\]

Let
\[
\gamma^{(k)} := \alpha \left( \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)}) - \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)}) \right)
\]
and note that using Lemma 2, we have
\[
\|\gamma^{(k)}\| \leq \alpha L_x e_y.
\]
Continuing the derivation in (36), we obtain
\[
\|x^{(k)} - x^{(s,k)} - \alpha \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
\leq \|x^{(k)} - x^{(s,k)} - \alpha \nabla_x L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
+ \frac{2}{\alpha^2} \left( g + \frac{M_\gamma e_y}{2} \right)\| \lambda^{(k+1)} - \lambda \|
\leq \|x^{(k)} - x^{(s,k)}\|^2 + \|x^{(s,k)} - x^{(s,k+1)}\|^2
+ 2\|x^{(s,k)} - x^{(s,k+1)}\|^2
+ \frac{2B + \alpha F_x}{\alpha} \|x^{(s,k)} - x^{(s,k+1)}\|^2
+ \alpha^2 g^2 \left[ \frac{4}{\alpha^2} + \alpha(2g + M_\gamma e_y) \right] M_\gamma e_y
\leq \|x^{(k)} - x^{(s,k)}\|^2 + \|x^{(s,k)} - x^{(s,k+1)}\|^2
+ 2\|x^{(s,k)} - x^{(s,k+1)}\|^2
+ \alpha^2 g^2 \left[ \frac{4}{\alpha^2} + \alpha(2g + M_\gamma e_y) \right] M_\gamma e_y
\leq \|x^{(k)} - x^{(s,k)}\|^2 + \|x^{(s,k)} - x^{(s,k+1)}\|^2
+ \frac{2B + \alpha F_x}{\alpha} \|x^{(s,k)} - x^{(s,k+1)}\|^2
+ \alpha^2 g^2 \left[ \frac{4}{\alpha^2} + \alpha(2g + M_\gamma e_y) \right] M_\gamma e_y
\]

where the first inequality holds by the Cauchy-Schwarz inequality, (37), and Assumption 2, and the last inequality holds by the convexity of $L^{(k)}(\cdot, \lambda)$. The first part of the lemma then follows by combining (35) and (38), and rearranging.

For the second part, for any $\lambda \in \Lambda_{\alpha,n}$, we have that
\[
\|\lambda^{(k+1)} - \lambda\|^2
\leq \|\lambda^{(k+1)} - \lambda + \alpha g^{(k)}(\tilde{x}^{(k)})\|^2
\leq \|\lambda^{(k+1)} - \lambda\|^2 + \alpha \|\nabla_{\lambda} L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
+ \alpha \|g^{(k)}(\tilde{x}^{(k)}) - \nabla_{\lambda} L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
\]
where
\[
\beta^{(k)} := \alpha \left( g^{(k)}(\tilde{x}^{(k)}) - \nabla_{\lambda} L^{(k)}(x^{(k)}, \lambda^{(k)}) \right)
\]
and noticing that, by Lemma 2,
\[
\|\beta^{(k)}\|_2 \leq \alpha M_g e_y,
\]
it follows that
\[
\|\lambda^{(k+1)} - \lambda\|^2
\leq \|\lambda^{(k)} - \lambda + \alpha \nabla_{\lambda} L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
+ \alpha \|\nabla_{\lambda} L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
\leq \|\lambda^{(k)} - \lambda\|^2 + 2\alpha \|\nabla_{\lambda} L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
+ \alpha^2 g^2 \left[ \frac{4}{\alpha^2} + \alpha(2g + M_\gamma e_y) \right] M_\gamma e_y
\leq \|\lambda^{(k+1)} - \lambda\|^2 + 2\alpha \|\nabla_{\lambda} L^{(k)}(x^{(k)}, \lambda^{(k)})\|^2
+ \alpha^2 g^2 \left[ \frac{4}{\alpha^2} + \alpha(2g + M_\gamma e_y) \right] M_\gamma e_y
\]

where the second inequality holds by (42); and the equality holds by the linearity of $\mathcal{L}^{(k)}(x, \lambda)$ in $\lambda$. The second part of the lemma then follow by rearranging the obtained inequality. □

With these intermediate results on place, the proofs of Theorem 1 and Theorem 2 are provided next.

**Proof of Theorem 4** By using Lemma 3 we have that

$$\mathcal{L}^{(k)}(x^{(s,k)}, \lambda) - \mathcal{L}^{(k)}(x^{(s,k)}, \lambda)$$

$$\leq \left[ \|x^{(s,k)}\|^2 - \|x^{(s,k)}\|^2 \right] / (2\alpha) + \alpha F_{x}^2 / 2$$

$$+ \left[ (B + \alpha F_{x}) + \frac{\alpha}{2} L_{e} x_{y} \right] L_{x} x_{y}$$

$$+ \frac{\|\lambda^{(k)} - \lambda^{(k+1)} - \lambda^{(k)}\|^2 / (2\alpha) + \alpha g_{x}^2 / 2}{\alpha}$$

$$+ \left[ \frac{2}{\alpha} + \frac{\alpha}{2} \left( g + \frac{M_{e} y_{y}}{\alpha} \right) \right] M_{e} y_{y}$$

(40)

for any $\lambda \in \Lambda_{k}$. To show (40), we use $\lambda = 0$ and the fact that $|\lambda^{(k)}| g^{(k)}(y^{(k)}(x^{(s,k)})) \leq 0$ by the feasibility of $x^{(s,k)}$ for (P0)$_{(k)}$. Therefore, by (46), we have that

$$\mathcal{L}^{(k)}(x^{(k)}, 0) - \mathcal{L}^{(k)}(x^{(s,k)}, \lambda)$$

$$= h^{(k)}(x^{(k)}) - h^{(k)}(x^{(s,k)}) - |\lambda^{(k)}| g^{(k)}(y^{(k)}(x^{(s,k)}))$$

$$\geq h^{(k)}(x^{(k)}) - h^{(k)}(x^{(s,k)}).$$

(41)

By using this last inequality in (40), and summing (40) over $\ell = 1, \ldots, k$, we have

$$\frac{1}{k} \sum_{\ell = 1}^{k} \left( h^{(k)}(x^{(\ell)}) - h^{(k)}(x^{(s,\ell)}) \right)$$

$$\leq \frac{1}{2\alpha} \left( \|x^{(1)} - x^{(s,1)}\|^2 + \|\lambda^{(1)}\|^2 \right) + \frac{\alpha}{2} (F_{x}^2 + g_{x}^2)$$

$$+ \left[ (B + \alpha F_{x}) + \frac{\alpha}{2} L_{e} x_{y} \right] L_{x} x_{y}$$

$$+ \left[ \frac{2}{\alpha} + \frac{\alpha}{2} \left( g + \frac{M_{e} y_{y}}{\alpha} \right) \right] M_{e} y_{y}$$

$$+ \frac{\alpha F_{x}^2}{2} = F + G / \alpha \alpha^{\kappa},$$

and hence

$$\alpha F_{x}^2 = \alpha (F_{x}^2 + 2FG / \alpha \alpha + G^2 / \alpha^2 \alpha)$$

$$= \alpha F_{x}^2 + 2FG \alpha^{-\kappa} + G^2 \alpha^{2-2\kappa}.$$ 

Using this, the fact that $\|x^{(1)} - x^{(s,1)}\| \leq 2B$, and assuming (without loss of generality) that $\lambda^{(1)} = 0$, we complete the proof of (42) and (43).

**Proof of Theorem 5** To prove (44), for a given $j = 1, \ldots, M$, consider the $j$-th component of $g^{(k)}(y^{(k)}(x^{(s,k)}))$, and let $\lambda_{j}$ be a vector in $\mathbb{R}^{M}$ with all zero components apart from the $j$-th component which equals $1 / \alpha$. Note that $\lambda_{j} \in \Lambda_{k}$. By construction, and $\lambda_{j}^{2} g^{(k)}(y^{(k)}(x^{(s,k)})) = g^{(k)}(y^{(k)}(x^{(s,k)})) / \alpha$. Therefore,

$$\mathcal{L}^{(k)}(x^{(k)}, \lambda_{j}) - \mathcal{L}^{(k)}(x^{(s,k)}, \lambda_{j})$$

$$\geq h^{(k)}(x^{(k)}) - h^{(k)}(x^{(s,k)}) + \frac{1}{\alpha} g^{(k)}(y^{(k)}(x^{(s,k)})).$$

Now, by convexity of $h$ and Assumption 2 we have that

$$h^{(k)}(x^{(s,k)}) - h^{(k)}(x^{(k)}) \leq \left( \nabla h^{(k)}(x^{(s,k)}) \right)^{T} (x^{(s,k)} - x^{(k)})$$

$$\leq \left\| \nabla h^{(k)}(x^{(s,k)}) \right\| \left\| x^{(s,k)} - x^{(k)} \right\|$$

$$\leq 2FB$$

(43)

Thus, letting $B(\infty)(\alpha, \kappa)$ denote the asymptotic bound of (42), and using similar derivation, completes the proof of the theorem. □

**IV. Tracking of Time-Varying Saddle Points**

Let $p > 0$ and $d > 0$, and consider re-writing the algorithmic steps (12) in the following compact form:

$$z^{(k+1)} = \text{Proj}_{\chi^{(k)} \times D^{(k)}} \left\{ z^{(k)} - p \phi(z^{(k)}) \right\},$$

(44)

with the time-varying map $\phi : \chi^{(k)} \times D^{(k)} \to \mathbb{R}^{n} \times \mathbb{R}^{m}$ defined as

$$\phi^{(k)} : z \mapsto \left[ \begin{array}{c} \nabla_{x} \mathcal{L}^{(k)}(z) \\ -g^{(k)}(y^{(k)}(x)) + d\lambda \end{array} \right],$$

(45)

and where, similarly to (22), $\nabla_{x} \mathcal{L}^{(k)}_{p,d}(z)$ is the approximate gradient of the regularized Lagrangian function calculated as:

$$\nabla_{x} \mathcal{L}^{(k)}_{p,d}(z) = \nabla_{x} \mathcal{L}^{(k)}(z) - p\tilde{x}$$

(46)

Similar to the previous section, let $\nabla_{x} \mathcal{L}^{(k)}_{p,d}(z) = \nabla_{x} \mathcal{L}^{(k)}(z) - p\tilde{x}$ be the gradient of the regularized Lagrangian evaluated at $z := (x, \lambda)$ and at the the synthetic output $y^{(k)}(x)$. Using $\nabla_{x} \mathcal{L}^{(k)}_{p,d}(z)$, let $\phi^{(k)} : \chi^{(k)} \times D^{(k)} \to \mathbb{R}^{n} \times \mathbb{R}^{m}$ be the counterpart of $\phi^{(k)}$ when the model $y^{(k)}(x)$ is utilized; that is,

$$\phi^{(k)} : z \mapsto \left[ \begin{array}{c} \nabla_{x} \mathcal{L}^{(k)}_{p,d}(z) \\ -g^{(k)}(y^{(k)}(x)) + d\lambda \end{array} \right],$$

(47)

Replacing $\phi^{(k)}$ with $\phi^{(k)}$ in (44) yields a feed-forward online algorithm, as discussed in Section 4. The following lemma bounds the difference between the outputs of the maps $\phi^{(k)}$ and $\phi^{(k)}$.

**Lemma 4.** Assumptions 2, 3, 4 and 5 the perturbation in the map $\phi^{(k)}$ can be bounded as:

$$\left\| \phi^{(k)}(z^{(k)}) - \phi^{(k)}(z^{(k)}) \right\|_{2} \leq e_{p} + e_{d}$$

(48)

where

$$e_{p} \leq L_{0} + M_{\lambda} M_{L} \max_{j=1, \ldots, M} \{ L_{a_{j}} \} \left\| C \right\|_{2} e_{y}$$

(49)

$$e_{d} \leq M_{g} e_{y},$$

(50)

with $M_{\lambda} := \sup_{k \geq 1} \max_{\lambda \in D^{(k)}} \| \lambda \|_{1}$.

**Proof.** Notice first that the left hand side of (48) can be written as $\| e_{p}^{(k)} \|^{2} + \| e_{d}^{(k)} \|^{2}$, where

$$e_{p}^{(k)} := \nabla_{x} \mathcal{L}^{(k)}_{p,d}(z^{(k)}) - \nabla_{x} \mathcal{L}^{(k)}_{p,d}(z^{(k)})$$

(51)

$$e_{d}^{(k)} := g^{(k)}(y^{(k)}(x^{(k)})) - g^{(k)}(y^{(k)}).$$

(52)
Regarding (52), from [19] and Assumption [8] it follows that
\[ \|e_d(k)\|_2 \leq M_\gamma \|y(k)(x^{(k)}) - \tilde{y}(k)\|_2 \leq M_\gamma e_y \]  
for all \( k \in \mathbb{N} \). Regarding (51), use the triangle inequality to obtain \( \|e_p(k)\|_2 \leq \|e_p(1,1)\|_2 + \|e_p(2,2)\|_2 \), with:
\[
e_p(k) := \sum_{j=1}^M \lambda_j^k \mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k))) \tag{54}
\]
\[
e_p(k) := \sum_{j=1}^M \lambda_j^k \mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k))) \tag{55}
\]
The first term can be bounded as
\[
\|e_p(k)\|_2 \leq \|\|\mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k)))\|_2 \leq C_0 \|e_y\|_2 \tag{56a}
\]
\[
\|e_p(k)\|_2 \leq \|\|\mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k)))\|_2 \leq C_0 \|e_y\|_2 \tag{56b}
\]
\[
\|e_p(k)\|_2 \leq \|\|\mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k)))\|_2 \leq C_0 \|e_y\|_2 \tag{56c}
\]
The norm of \( e_p(k) \) can be bounded as follows:
\[
\|e_p(k)\|_2 \leq \|\|\mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k)))\|_2 \leq C_0 \|e_y\|_2 \tag{57a}
\]
\[
\|e_p(k)\|_2 \leq \|\|\mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k)))\|_2 \leq C_0 \|e_y\|_2 \tag{57b}
\]
\[
\|e_p(k)\|_2 \leq \|\|\mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k)))\|_2 \leq C_0 \|e_y\|_2 \tag{57c}
\]
\[
\|e_p(k)\|_2 \leq \|\|\mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k)))\|_2 \leq C_0 \|e_y\|_2 \tag{57d}
\]
\[
\|e_p(k)\|_2 \leq \|\|\mathbf{C}^T (\nabla x g_j(y(k)(x^{(k)})) - \nabla x g_j(\tilde{y}(k)))\|_2 \leq C_0 \|e_y\|_2 \tag{57e}
\]
Using the definition of \( M_\lambda \), the result follows.

Notice that (48) naturally bounds the errors in the computation of gradients \([29], [30]\) that one commits by “closing the loop”; i.e., by replacing the model \( y^{(k)}(x^{(k)}) \) with the measurements \( \tilde{y}^{(k)} \).

Before stating the main convergence result, the following assumption is imposed on the mapping \( \phi^{(k)}(z) \).

**Assumption 6.** For every \( k \in \mathbb{N} \), the map \( \phi^{(k)}(z) \) is strongly monotone over \( \mathcal{X}^{(k)} \times D^{(k)} \) with constant \( \eta_\phi := \min\{p, d\} \) and Lipschitz over \( \mathcal{X}^{(k)} \times D^{(k)} \) with coefficient \( L_\phi \) given by:
\[
L_\phi := \sqrt{(L + p + M_d + \xi_\lambda L_g)^2 + (M_d + d)^2} \tag{60}
\]

where \( L_g := \sqrt{\sum_{m=1}^{M_1} L_{2,g_m}^2} \) and \( \xi_\lambda := \sup_{k \in \mathbb{N}} M_{i-1,\ldots,i} \{L_{2,g_i}\} e_y \).

Theorem 3 provides a bound on \( \|z^{(k)} - z^{(*,k)}\|_2 \) per each time instant \( k \in \mathbb{N} \). Notice that \( \|\phi(z^{(k)}) - \phi(z^{(k)})\|_2 \) on the right-hand-side of (58) can be bounded as shown in Lemma 4; the next assumption is introduced and utilized to establish asymptotic properties of \( \|z^{(k)} - z^{(*,k)}\|_2 \).

**Assumption 7.** There exists a scalar \( \bar{\sigma} < +\infty \) such that \( \sup_{k \geq 1} \bar{\sigma}(\kappa) \leq \bar{\sigma} \).

Notice that Assumption 7 is always satisfied in practice since it is assumed that the sets \( \mathcal{X}^{(k)} \) and \( D^{(k)} \) are compact uniformly in time (and, therefore, optimal solutions are never unbounded) and the solution set is not empty. For \( s \to 0, \bar{\sigma} \) is in fact an upper bound on the norm of the gradient of the optimal trajectory \( \{z^{(*,k)}\}_{k \in \mathbb{N}} \). The next result will establish Q-linear convergence.

**Corollary 1.** Suppose that Assumption 7 holds and \( c(\alpha) < 1 \). Then, the sequence \( \{z^{(k)}\} \) converges Q-linearly to \( \{z^{(*,k)}\} \) up to an asymptotic error bound given by:
\[
\limsup_{k \to \infty} \|z^{(k)} - z^{(*,k)}\|_2 \leq \frac{\alpha \sqrt{\bar{c}_p^2 + \bar{c}_d^2 + \bar{\sigma}}}{1 - c(\alpha)} \tag{61}
\]

The proofs of Theorem 3 and Corollary 1 follow steps that are similar to the ones outlined in [2]: a summary of the steps as well as modifications relative to [2] are provided next for completeness.
Proof of Theorem 3. Start from the the following equation:
\[
\|z^{(k)} - z^{(s,k-1)}\|_2 = \|\text{Proj}_{z^{(k)} \times D^{(k)}} \{z^{(k-1)} - \hat{\phi}^{(k)}(z^{(k-1)})\} - z^{(s,k-1)}\|_2. \tag{62}
\]
Noticing that \(z^{(s,k-1)}\) satisfies a fixed-point equation, leveraging the non-expansiveness property of the projection operator, the following inequality can be considered:
\[
\|z^{(k)} - z^{(s,k-1)}\|_2 \leq \|z^{(k-1)} - \alpha\hat{\phi}^{(k-1)}(z^{(k-1)})\|_2 - \|z^{(s,k-1)} + \alpha\hat{\phi}^{(k-1)}(z^{(s,k-1)})\|_2. \tag{63}
\]
Adding and subtracting \(\hat{\phi}^{(k-1)}(z^{(k-1)})\) on the right-hand-side of (63), and using the triangle inequality, it follows that (63) can be further bounded as:
\[
\|z^{(k)} - z^{(s,k-1)}\|_2 \leq \alpha \|\phi^{(k-1)}(z^{(k-1)}) - \hat{\phi}^{(k-1)}(z^{(k-1)})\|_2 + \|z^{(k-1)} - \alpha \Phi^{(k-1)}(z^{(k-1)})\|_2 - \alpha \Phi^{(k-1)}(z^{(s,k-1)})\|_2. \tag{64}
\]

Following [2], using the results of Lemma 5, the second term on the right-hand-side of (64) can be bounded with the term \(c(\alpha)\|z^{(k-1)} - z^{(s,k-1)}\|_2\); therefore,
\[
\|z^{(k)} - z^{(s,k-1)}\|_2 \leq \alpha \|\phi^{(k-1)}(z^{(k-1)}) - \hat{\phi}^{(k-1)}(z^{(k-1)})\|_2 + c(\alpha)\|z^{(k-1)} - z^{(s,k-1)}\|_2. \tag{65}
\]
Consider now bounding \(\|z^{(k)} - z^{(s,k)}\|_2\) as follows:
\[
\|z^{(k)} - z^{(s,k)}\|_2 = \|z^{(k)} - z^{(k-1)} + z^{(s,k-1)} - z^{(s,k)}\|_2 \\
\leq \|z^{(k-1)} - z^{(s,k-1)}\|_2 + \|z^{(k)} - z^{(s,k-1)}\|_2 \\
\leq \alpha \|\phi^{(k-1)}(z^{(k-1)}) - \hat{\phi}^{(k-1)}(z^{(k-1)})\|_2 \\
+ c(\alpha)\|z^{(k-1)} - z^{(s,k-1)}\|_2. \tag{66}
\]
By recursively applying (67), the result of Theorem 3 follows.

To show (61), utilize the results of Lemma 4 to bound \(\|\phi^{(k)}(z^{(k-1)}) - \hat{\phi}^{(k)}(z^{(k-1)})\|_2\) and leverage Assumption 7. Then, (67) can be bounded as:
\[
\|z^{(k)} - z^{(s,k)}\|_2 \\
\leq c(\alpha)\|z^{(k-1)} - z^{(s,k-1)}\|_2 + \bar{\sigma} + \alpha \sqrt{\epsilon_p^2 + \epsilon_d^2}. \tag{68}
\]
since \(c(\alpha) < 1\), (68) represents a contraction. The result (61) can then be obtained via the geometric series sum formula. \(\Box\)

The coefficient \(c(\alpha)\) is less then one when \(\alpha < 2\bar{\epsilon}/L^2\).
When no measurement errors are present, \(\epsilon_p = \epsilon_d = 0\) and (61) provides a result for feed-forward online algorithms (similar to e.g., [26, 28]). When \(\bar{\sigma} = 0\), then the underlying optimization problem is static and the algorithm converges to the solution of the static optimization problem (11). Finally, notice that the result (61) can also be interpreted as input-to-state stability, where the optimal trajectory \(\{z^{(s,k)}\}\) of the time-varying minimax problem (11) is taken as a reference.

V. CONCLUSION

This paper leveraged a time-varying convex optimization formalism to model optimal operational trajectories of systems or network of systems, and developed feedback-based online algorithms based on primal-dual projected-gradient methods. In the proposed algorithms, the gradient steps were modified to accommodate measurements from the network system. When the design of the algorithm is based on the time-varying Lagrangian, the paper characterized the performance of the proposed via a dynamic regret analysis. When a regularized Lagrangian is utilized, results in terms of Q-linear convergence are provided, at the cost of tracking an approximate KKT trajectory.

Extending the proposed methodology to time-varying non-convex problems is the subject of current research efforts. Future efforts will also look at characterizing the performance of the propose method when implemented in an asynchronous fashion.

REFERENCES

[1] A. Simonetto and G. Leus, “Distributed asynchronous time-varying constrained optimization,” in 48th Asilomar Conference on Signals, Systems and Computers, Nov 2014, pp. 2142–2146.

[2] E. Dall’Anese and A. Simonetto, “Optimal power flow pursuit,” IEEE Trans. on Smart Grid, 2016.

[3] S. Rahili and W. Ren, “Distributed Convex Optimization for Continuous-Time Dynamics with Time-Varying Cost Functions,” IEEE Transactions on Automatic Control, 2016.

[4] M. Fazlyab, C. Nowzari, G. J. Pappas, A. Ribeiro, and V. M. Preciado, “Self-Triggered Time-Varying Convex Optimization,” in Proceedings of the 55th IEEE Conference on Decision and Control, Las Vegas, NV, US, December 2016, pp. 3090 – 3097.

[5] M. J. Neely and H. Yu, “Online convex optimization with time-varying constraints,” 2017, arXiv preprint:1702.04783.

[6] F. Bullo, Lectures on Network Systems, 2018, with contributions by J. Cortes, F. Dorfler, and S. Martinez. [Online] Available at: http://motion.me.ucsb.edu/book-ins.

[7] K. J. Arrow, L. Hurwicz, and H. Uzawa, Studies in Linear and Nonlinear Programming. Stanford, CA: Stanford University Press, 1958.

[8] A. Cherukuri, B. Gharesifard, and J. Cortes, “Saddle-point dynamics: conditions for asymptotic stability of saddle points,” SIAM J. Control Optim., vol. 55, no. 1, pp. 486–511, 2017.

[9] M. Zinkevich, “Online convex programming and generalized infinitesimal gradient ascent,” in Proceedings of the Twentieth International Conference on Machine Learning, (ICML 2003), August 21-24, 2003, Washington, DC, USA, 2003, pp. 928–936.

[10] J. Koshal, A. Nedić, and U. Y. Shanbhag, “Multitester optimization: Distributed algorithms and error analysis,” SIAM J. on Optimization, vol. 21, no. 3, pp. 1046–1081, 2011.

[11] M. B. Khuzani and N. Li, “Distributed regularized primal-dual method: convergence analysis and trade-offs,” [Online] Available at https://arxiv.org/abs/1609.08262.

[12] Y. Tang, K. Dvijotham, and S. Low, “Real-time optimal power flow,” IEEE Trans. on Smart Grid, vol. 8, no. 6, pp. 2963–2973, 2017.

[13] X. Zhou, E. Dall’Anese, L. Chen, and A. Simonetto, “An incentive-based online optimization framework for distribution grids,” IEEE Trans. on Automatic Control, 2017, to appear. [Online] Available at: https://arxiv.org/abs/1705.04182.

[14] A. Jokić, M. Lazar, and P. Van den Bosch, “On constrained steady-state regulation: Dynamic KKT controllers,” IEEE Trans. Auto. Contr., vol. 54, no. 9, pp. 2250–2254, Sep. 2009.

[15] S. Bolognani, R. Carli, G. Cavraro, and S. Zampieri, “Distributed reactive power feedback control for voltage regulation and loss minimization,” IEEE Trans. on Automatic Control, vol. 60, no. 4, pp. 966–981, Apr. 2015.

[16] K. Hirata, J. P. Hespanha, and K. Uchida, “Real-time pricing leading to optimal operation under distributed decision makings,” in Proc. of American Control Conf., Portland, OR, June 2014.
[17] A. Hauswirth, S. Bolognani, G. Hug, and F. Dorfler, “Projected gradient descent on Riemannian manifolds with applications to online power system optimization,” in 54th Annual Allerton Conference on Communication, Control, and Computing, Sept 2016, pp. 225–232.

[18] A. Bernstein, L. Reyes Chamorro, J.-Y. Le Boudec, and M. Paolone, “A compositional method for real-time control of active distribution networks with explicit power set points. part I: Framework,” Electric Power Systems Research, vol. 125, no. August, pp. 254–264, 2015.

[19] S. Hosseini, A. Chapman, and M. Mesbah, “Online distributed ADMM via dual averaging,” in 53rd IEEE Conference on Decision and Control, Dec 2014, pp. 904–909.

[20] A. Koppel, F. J. Jakubiec, and A. Ribeiro, “A saddle point algorithm for networked online convex optimization,” IEEE Transactions on Signal Processing, vol. 63, no. 19, pp. 5149–5164, Oct 2015.

[21] S. Lee and M. M. Zavlanos, “On the sublinear regret of distributed primal-dual algorithms for online constrained optimization,” 2017.

[22] T. Chen, Q. Ling, and G. B. Giannakis, “An online convex optimization approach to proactive network resource allocation,” IEEE Transactions on Signal Processing, vol. 65, no. 24, pp. 6350–6364, Dec 2017.

[23] T. Chen and G. B. Giannakis, “Bandit convex optimization for scalable and dynamic IoT management,” 2017, [Online] Available at: https://arxiv.org/abs/1707.09060.

[24] A. Nedić and A. Ozdaglar, “Subgradient methods for saddle-point problems,” J. of Optimization Theory and Applications, vol. 142, no. 1, pp. 205–228, 2009.

[25] I. Necoara and V. Nedelcu, “On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems,” Automatica, vol. 55, pp. 209–216, 2015.

[26] A. Simonetto and G. Leus, “Double smoothing for time-varying distributed multiuser optimization,” in IEEE Global Conf. on Signal and Information Processing, Dec. 2014.

[27] Q. Ling and A. Ribeiro, “Decentralized dynamic optimization through the alternating direction method of multipliers,” IEEE Trans. on Signal Processing, vol. 62, no. 5, pp. 1185–1197, Mar. 2014.

[28] A. Simonetto, “Time-varying convex optimization via time-varying averaged operators,” 2017, [Online] Available at: http://arxiv.org/abs/1704.07338.

[29] D. P. Bertsekas and J. N. Tsitsiklis, “Gradient convergence in gradient methods with errors,” SIAM J. on Optimization, vol. 10, no. 3, pp. 627–642, July 1999.

[30] T. Larsson, M. Patriksson, and A.-B. Strömberg, “On the convergence of conditional epsilon-subgradient methods for convex programs and convex-concave saddle-point problems,” European J. of Operational Research, vol. 151, no. 3, pp. 461–473, 2003.

[31] I. Necoara and V. Nedelcu, “Rate analysis of inexact dual first-order methods application to dual decomposition,” IEEE Trans. on Automatic Control, vol. 59, no. 5, pp. 1232–1243, 2014.

[32] J. Wang and N. Elia, “A control perspective for centralized and distributed convex optimization,” in Proc. of 50th IEEE Conf. on Decision and Control, Orlando, FL, Dec. 2011.

[33] F. D. Brunner, H.-B. Durr, and C. Ebenbauer, “Feedback design for multi-agent systems: A saddle point approach,” in Proc. of 51st IEEE Conf. on Decision and Control, Maui, HI, Dec 2012, pp. 3783–3789.

[34] N. Li, L. Chen, C. Zhao, and S. H. Low, “Connecting automatic generation control and economic dispatch from an optimization view,” in Proc. of American Control Conf., Portland, OR, June 2014.

[35] X. Zhang and A. Papachristodoulou, “Distributed dynamic feedback control for smart power networks with tree topology,” in Proc. of American Control Conf., Portland, OR, June 2014.

[36] A. Ribeiro, N. D. Sidiropoulos, and G. B. Giannakis, “Optimal distributed stochastic routing algorithms for wireless multihop networks,” IEEE Trans. on Wireless Communications, vol. 7, no. 11, pp. 4261–4272, Nov. 2008.

[37] A. Ephremides, “Energy concerns in wireless networks,” IEEE Wireless Communications, vol. 9, no. 4, pp. 48–59, Aug. 2002.

[38] J. Monteil, N. O’Hara, V. Cahill, and M. Bourroche, “Real-time estimation of drivers’ behavior,” in 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Sept 2015, pp. 2046–2052.

[39] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, “Dynamical model of traffic congestion and numerical simulation,” Physical review E, vol. 51, no. 2, p. 1035, 1995.

[40] M. Schütz, A. Campisano, H. Colas, P. Vanrolleghem, and W. Schilling, “Real-time control of urban water systems,” in Proc. of Intl. conf. on Pumps, Electromechanical Devices and Systems Applied to Urban Water Management, 2003, pp. 1–19.