Numerical Optimisation of Time-Varying Strongly Convex Functions Subject to Time-Varying Constraints

Daniel D. Selvaratnam[1], Iman Shames[1], Jonathan H. Manton[1], Mohammad Zamani[2]

***Abstract*— This paper analyses the performance of projected gradient descent on optimisation problems with cost functions and constraints that vary in discrete time. Specifically, strongly convex cost functions with Lipschitz gradient, and a sequence of convex constraints are assumed. Error bounds and sub-optimality bounds are derived for a variety of cases, which show convergence to a steady-state. Conditions on the constraint sequence are also presented for guaranteeing finite-time feasibility, and for bounding the distance between successive minimisers. Numerical examples are then presented to validate the analytical results.**

I. Introduction

Recent literature on numerical optimisation has explored problems involving cost functions and possibly constraints that change discretely with time. These arise whenever the parameters in a given optimisation problem are determined by a time-varying environment. Objectives or constraints may also change in response to updated information, such as the arrival of new measurements [6]. This behaviour is therefore relevant to a variety of problems in control, signal processing and robotics [6], [13], [18], [22]. Such time-varying problems can be considered a sequence of optimisation problems. Assuming that every cost function and feasible set in this sequence is made available to the solver, one approach would be to solve each individual problem completely. This may not be tractable, depending on the rate at which new functions arrive. It is also unnecessary if the primary objective is to generate a sequence of iterates that 'track' the time-varying optimal points, or to achieve asymptotically low cost. Assuming there are bounds on the variation between consecutive cost functions and constraints, information from the current problem can be exploited to optimise its successor. A more efficient alternative, therefore, is to only solve each problem partially by limiting the number of iterations per cost function. Such approaches have been termed *running* methods in [13]. Here, we consider the most extreme version of this, in which only a single iteration is performed per cost function. We focus on the performance of projected gradient descent iterations, applied to a sequence of smooth, strongly convex cost functions, subject to convex constraints. Bounds on sub-optimality and distance to the minima (henceforth referred to as *tracking error*) are derived and verified by means of numerical examples.

We now present a brief review of existing literature on time-varying optimisation. Gradient descent on a sequence of smooth, strongly convex functions is investigated in [12], with bounds on tracking error derived for unconstrained problems. Error bounds are also presented for cost functions that vary in continuous time, with the gradient descent iterates replaced by a control law. A general framework for time-varying convex optimisation problems (with time-varying constraints) is proposed in [13], based on the theory of averaged operators. The authors develop error bounds on a variant of the Mann-Krasnosel'skii iterations with time-varying operators. These bounds are then used to prove the convergence of a variety of running methods, including projected gradient descent, proximal-point, forward-backward splitting, dual ascent, dual decomposition, and ADMM. The tracking error bounds for (projected) gradient descent in [13] are stated under slightly different assumptions to [12], which requires the difference between consecutive cost function gradients to be bounded over the whole of $\mathbb{R}^n$. The latter is quite restrictive, given that strong convexity and an unconstrained domain are both assumed. In [13], a bounded distance between consecutive optimisers is assumed instead, with the results remaining valid under time-varying constraints. While this is more general, in practice it can be difficult to establish a useful bound on the distance between consecutive optimisers when time-varying constraints are involved (e.g. [3]). In our work, we derive such bounds under appropriate restrictions on the change between consecutive constraints.

Time-varying convex optimisation problems are studied by the machine learning community in the context of Online Convex Optimisation (OCO) [5], [23]. These differ from the aforementioned works in that OCO is concerned with minimising the integrated cost rather than the asymptotic cost. Compact constraints are a standard assumption in OCO literature that is not required in more general formulations. Algorithmic performance is evaluated in terms of the *regret*, defined as the integrated difference between the observed cost and the cost at the best fixed decision variable. *Dynamic regret* is a generalisation which compares the observed cost with the time-varying optimal cost. OCO results are typically presented in terms of regret bounds, the goal being to achieve sub-linear regret. Dynamic regret is more relevant to the wider literature and to our own work, because sub-linear dynamic regret implies asymptotically optimal cost. Dynamic regret bounds for projected gradient descent are derived in

[1] Melbourne Information Decision and Autonomous Systems Laboratory, Dept. Electrical and Electronic Engineering, University of Melbourne. dselvaratnam@student.unimelb.edu.au, {ishames, jmanton}@unimelb.edu.au

[2] Land Division, Defence Science and Technology Group. mohammad.zamani@dst.defence.gov.au

This research is supported by DST Group under Collaborative Research Agreements MYIP #6874, MYIP #5923, and by the Defence Science Institute as an initiative of the State Government of Victoria.

[10] under the assumptions of strongly convex cost functions, and fixed, compact constraints. Special cases of time-varying constraints are considered in [11], along with possibly non-smooth cost functions, however only static regret bounds are presented. In contrast, our work considers general time-varying (convex) constraints that need not be compact.

We now mention literature on time-varying optimisation that considers a different class of iterations to projected gradient descent. In particular, [4], [9], [16] focus on Newton-type methods, with [9] discussing optimisation problems on manifolds with changing co-ordinate maps. Generalized equations are used in [21] to develop an augmented-Lagrangian method. Several other works have studied time-varying cost functions in the context of distributed optimisation over a network. The distributed structure of the solutions precludes the use of straightforward gradient descent. Of these, [17], [20] both treat the entire problem in continuous time, focusing on quadratic costs and deriving control laws which track the optimum. Others present iterative solutions using a variety of methods such as consensus-based algorithms in [2], [14], and distributed gradient descent in [19]. Duality is exploited by [6], [15], [22] and [7], which develops running ADMM.

Our own work extends the projected gradient descent results in [12] and [13] by exploring the consequences of time-varying constraints in greater detail. In addition to bounding the distance between consecutive minimisers, further conditions are stated which guarantee finite-time feasibility. We also present sub-optimality bounds, along with a more detailed analysis of the error than is available in the above works. The remainder of this paper is structured as follows. The problem and its basic assumptions are stated precisely in Section II. General error bounds are presented in Section III, along with sub-optimality bounds for unconstrained problems. Section IV specifically tackles issues relating to constraints. Two numerical examples are then provided in Section V, which validate the theoretical bounds, and illustrate the assumptions that were made.

II. Problem Formulation

Let $\mathbb{N} := \{1, 2, ...\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.
We consider a family of continuous functions $\mathcal{F} := \{f_k : \mathbb{R}^n \to \mathbb{R} \mid k \in \mathbb{N}_0\}$, and a family of closed, non-empty sets $\mathcal{X} := \{X_k \subset \mathbb{R}^n \mid k \in \mathbb{N}_0\}$ capturing the possible time-varying constraints. We are interested in generating solutions to the following sequence of optimisation problems

$$\min_{x \in X_k} f_k(x), \tag{1}$$

using iterates of the form

$$x_{k+1} = G_k(x_k), \tag{2}$$

where the operator $G_k : \mathbb{R}^n \to \mathbb{R}^n$ is restricted to using first-order information and knowledge of the feasible set X_k at the current time step. Note that this formulation reduces to a standard time-invariant optimisation problem when $\mathcal{F} = \{f_0\}$ and $\mathcal{X} = \{X_0\}$.

In this paper, we restrict attention to families of smooth, strongly convex functions with Lipschitz gradient, and convex constraints. These are stated explicitly in the assumptions below.

Assumption 1 (Smoothness): Every $f \in \mathcal{F}$ is twice continuously differentiable, and

$$\exists L > 0,\ \forall f \in \mathcal{F},\ \forall x, y \in \mathbb{R}^n,\ \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

Assumption 2 (Uniform strong convexity): There exists $\sigma \in (0, L]$ such that

$$\forall f \in \mathcal{F},\ \forall x \in \mathbb{R}^n,\ \nabla^2 f(x) \succeq \sigma I.$$

Assumption 3 (Convex constraints): Every $X \in \mathcal{X}$ is closed, convex and non-empty.
We employ projected gradient descent to solve this problem. This pertains to a specific choice of G_k that will be defined in the next section. Further assumptions will be made as required.

III. General Error Bounds

Let $P_X : \mathbb{R}^n \to X$,

$$P_X(y) := \arg\min_{x \in X} \|x - y\|^2,$$

denote the projection operator, where $X \subset \mathbb{R}^n$ is a closed, convex set. The Projection Theorem guarantees the existence and uniqueness of $P_X(y)$. Projected gradient descent iterates evolve according to (2), where

$$G_k(x) := P_{X_k}\left(x - \alpha_k \nabla f_k(x)\right), \tag{3}$$

and $\alpha_k > 0$ is a sequence of step sizes.

Remark 3.1: Gradient descent on the unconstrained problem can be recovered by choosing $\mathcal{X} = \{\mathbb{R}^n\}$.

Under Assumption 2, f_k is guaranteed to have a unique global minimiser over X_k, which we define as

$$x_k^\star := \arg\min\{f_k(x) \mid x \in X_k\}.$$

We will consider the evolution of the *tracking error* $e_k := \|x_k - x_k^\star\|$, and the distance

$$\bar{e}_k := \|x_{k+1} - x_k^\star\|,$$

which we refer to as the *estimation error*, because x_{k+1} can be considered an estimate of $x_k^\star$ using all the information available up to time k. In [1, Proposition 6.1.8], it is established that, under Assumptions 1 – 3, G_k is a contraction map for $\alpha_k < \frac{2}{L}$. Specifically,

$$\forall x, y \in \mathbb{R}^n,\ \|G_k(x) - G_k(y)\| \leq \rho_k\|x - y\|, \tag{4}$$

where

$$\rho_k := \max\{|1 - \alpha_k L|, |1 - \alpha_k \sigma|\}. \tag{5}$$

Given the sequence α_k, we will also find it convenient to define

$$\rho := \sup\{\rho_k \mid k \in \mathbb{N}_0\}.$$

It is well known that $x_k^\star$ is a fixed point of G_k, and we present a simple proof of this below.

Lemma 3.1 (Fixed point): Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and continuously differentiable. Let $X \subset \mathbb{R}^n$ be closed, convex and non-empty. Let $\alpha > 0$, and define

$$G(x) := P_X(x - \alpha \nabla f(x)).$$

Then, $G(x^\star) = x^\star$ for any $x^\star \in \arg\min\{f(x) \mid x \in X\}$.

Proof: By definition of $x^\star$,

$$\forall x \in X,\ \nabla f(x^\star)^\top (x - x^\star) \geq 0. \tag{6}$$

Now the Projection Theorem [8, Theorem 1 of Section 3.12] implies

$$\forall x \in X,\ [x^\star - \alpha \nabla f(x^\star) - G(x^\star)]^\top [x - G(x^\star)] \leq 0.$$

Substituting $x = x^\star \in X$ into the above yields

$$\|x^\star - G(x^\star)\|^2 + \alpha \nabla f(x^\star)^\top [G(x^\star) - x^\star] \leq 0.$$

Referring to (6), both terms on the LHS are non-negative, and therefore both must vanish. ∎

Substituting $x = x_k$ and $y = x_k^\star$ into (4) then yields the inequality

$$\bar{e}_k \leq \rho_k e_k, \tag{7}$$

which we use to bound the tracking error evolution. For notational convenience, define $\prod_{j=N}^{N-1} \rho_j := 1$ for any $N \in \mathbb{N}$.

Theorem 3.2 (Tracking error dynamics): Let $\mathcal{F}$, $\mathcal{X}$ satisfy Assumptions 1 – 3. Let $x_0 \in \mathbb{R}^n$, and consider iterations of the form

$$x_{k+1} = P_{X_k}\left(x_k - \alpha_k \nabla f_k(x_k)\right), \tag{8}$$

where $0 < \alpha_k < \frac{2}{L}$ for all k. Then

$$\forall k \in \mathbb{N},\ e_{k+1} \leq \rho_k e_k + \|x_{k+1}^\star - x_k^\star\|. \tag{9}$$

Proof: Observe that

$$\begin{aligned} \|x_{k+1} - x_{k+1}^\star\| &= \|x_{k+1} - x_k^\star + x_k^\star - x_{k+1}^\star\| \\ &\leq \|x_{k+1} - x_k^\star\| + \|x_k^\star - x_{k+1}^\star\| \\ &\leq \rho_k \|x_k - x_k^\star\| + \|x_k^\star - x_{k+1}^\star\|, \end{aligned} \tag{10}$$

by application of (7). ∎

Corollary 3.3: For all $N \in \mathbb{N}$,

$$e_N \leq e_0 \prod_{j=0}^{N-1} \rho_j + \sum_{k=0}^{N-1} \|x_{k+1}^\star - x_k^\star\| \prod_{j=k+1}^{N-1} \rho_j. \tag{11}$$

Proof: Treating (10) as a dynamical system, we obtain

$$\forall N \in \mathbb{N},\ e_N \leq \Phi(N, 0) e_0 + \sum_{k=0}^{N-1} \Phi(N, k+1) \|x_{k+1}^\star - x_k^\star\|,$$

with state-transition operator $\Phi(N, k) := \prod_{j=k}^{N-1} \rho_j$. ∎

Remark 3.2: Theorem 3.2 reveals the tracking error is bounded by a discrete time dynamical system, where the shift in minimisers behaves as a disturbance input. Corollary 3.3 presents a closed-form expression for this bound.

If the shift in minimizers (i.e., the disturbance input) is bounded, this leads to a steady-state error bound under mild restrictions on the step sizes. We make the following assumption accordingly.

Assumption 4 (Bounded change in minimiser): For the pair $(\mathcal{F}, \mathcal{X})$, there exists $V \geq 0$ such that

$$\forall k \in \mathbb{N}_0,\ \|x_{k+1}^\star - x_k^\star\| \leq V.$$

Applying this leads directly to an alternative proof of [13, Corollary 7.1 (b)].

Corollary 3.4: Let the hypotheses of Theorem 3.2 be satisfied, along with Assumption 4. If there also exists $\epsilon > 0$ and $\bar{\alpha} < \frac{2}{L}$ such that $\epsilon \leq \alpha_k \leq \bar{\alpha}$ for all k, then

$$\limsup_{k \to \infty} e_k \leq \frac{V}{1 - \rho} \tag{12}$$

$$\limsup_{k \to \infty} \bar{e}_k \leq \frac{V\rho}{1 - \rho}. \tag{13}$$

Proof: Referring to (5), ρ_k only attains its maximum value of 1 over $\alpha_k \in [0, \frac{2}{L}]$ at $\alpha_k = 0$ and $\alpha_k = \frac{2}{L}$. Thus, for any $\epsilon \in (0, \frac{2}{L})$ and $\bar{\alpha} \in (\epsilon, \frac{2}{L})$, restricting $\alpha_k \in [\epsilon, \bar{\alpha}]$ for all k implies

$$\rho \leq \sup_{\alpha \in [\epsilon, \bar{\alpha}]} \left(\max\{|1 - \alpha L|, |1 - \alpha\sigma|\} \right) < 1.$$

Since ρ and V are upper bounds for ρ_k and $\|x_{k+1}^\star - x_k^\star\|$ respectively, we obtain

$$e_N \leq \rho^N e_0 + V \sum_{k=0}^{N-1} \rho^k$$

from (11). The result follows by taking limits, and applying (7). ∎

Remark 3.3: When Assumption 4 is satisfied, it is clear from (9), (12) and (13) that minimising ρ_k results in both the fastest rate of convergence of the bound, and its steady-state value. Referring to (5), this corresponds to a constant step size of $\alpha_k = \frac{2}{\sigma + L}$, which implies $\rho = \rho_k = \frac{L - \sigma}{L + \sigma}$ for all k. This yields bounds of $\frac{V(L+\sigma)}{2\sigma}$ and $\frac{V(L-\sigma)}{2\sigma}$ in (12) and (13) respectively.

A. Unconstrained sub-optimality bounds

We now turn our attention to the level of sub-optimality. Specifically, we derive bounds on

$$\phi_k := f_k(x_k) - f_k(x_k^\star) \tag{14}$$

$$\bar{\phi}_k := f_k(x_{k+1}) - f_k(x_k^\star), \tag{15}$$

which are analogous to the tracking and estimation errors. We refer to $\bar{\phi}_k$ as the *predicted sub-optimality*, which is relevant if the cost of action x_k is incurred before the function changes. Here, we consider bounds for unconstrained problems. The effect of constraints will be considered in Section IV-C. The guarantee that $\nabla f_k(x_k^\star) = 0$ when $X_k = \mathbb{R}^n$ leads to the inequalities

$$\phi_k \leq \tfrac{L}{2} e_k^2 \tag{16}$$

$$\bar{\phi}_k \leq \tfrac{L}{2} \bar{e}_k^2, \tag{17}$$

which are proved in Lemma A.1 under Assumption 1. These allow our tracking and estimation error bounds to be translated directly into sub-optimality bounds. In particular, they yield the steady-state sub-optimality bounds below.

Corollary 3.5 (Steady-state sub-optimality): Suppose $\mathcal{X} = \{\mathbb{R}^n\}$ and $\mathcal{F}$ satisfies Assumptions 1, 2 and 4. Let $x_0 \in \mathbb{R}^n$, and consider iterations of the form

$$x_{k+1} = x_k - \alpha \nabla f_k(x_k), \tag{18}$$

where $\alpha = \frac{2}{\sigma+L}$. Then

$$\limsup_{k\to\infty} \phi_k \leq \frac{LV^2(L+\sigma)^2}{8\sigma^2} \tag{19}$$

$$\limsup_{k\to\infty} \bar{\phi}_k \leq \frac{LV^2(L-\sigma)^2}{8\sigma^2}. \tag{20}$$

IV. Analysis of Constraints

The previous error bounds apply equally well to constrained and unconstrained problems. However, the presence of constraints does warrant extra consideration. As yet, projected gradient descent does not guarantee $x_k \in X_k$. Establishing bounds on the shift in optima to satisfy Assumption 4 is also more challenging when constraints are involved. Furthermore, the sub-optimality bounds of Section III-A no longer apply because the constrained minima need not be stationary points. In this section, we present sufficient conditions to address the issues of feasibility, to bound the shift between minimisers, and to bound sub-optimality in the presence of time-varying constraints.

A. Finite-time feasibility

While projected gradient descent only guarantees feasibility with respect to the previous constraints, we can guarantee eventual feasibility under an additional assumption.

Assumption 5 (Sufficient overlap): The pair $(\mathcal{F}, \mathcal{X})$ is such that

$$\forall k \in \mathbb{N}_0,\ \{x \in X_k \mid \|x - x_k^\star\| \leq R\} \subset X_{k+1},$$

for some $R > \frac{V\rho}{1-\rho}$.

Proposition 4.1 (Finite-time feasibility): Let Assumptions 1 – 5 be satisfied. Let $x_0 \in \mathbb{R}^n$, and consider iterations of the form

$$x_{k+1} = x_k - \alpha \nabla f_k(x_k),$$

where $\alpha \in (0, \frac{2}{L})$. Then,

$$\exists N \in \mathbb{N}_0,\ \forall k > N,\ x_k \in X_k.$$

Proof: For any $R > \frac{V\rho}{1-\rho}$, (13) implies there exists $N \in \mathbb{N}_0$ such that $\|x_{k+1} - x_k^\star\| \leq R$ for all $k \geq N$. Use of the projection operator P_{X_k} in (8) ensures $x_{k+1} \in X_k$, and Assumption 5 then guarantees $x_{k+1} \in X_{k+1}$ for all $k \geq N$. ■

B. Bounds on the shift between minimisers

Here, we present alternative conditions that guarantee Assumption 4. These conditions impose restrictions on changes in the cost function gradients and the constraints between consecutive time steps.

Lemma 4.2: Let $X_1, X_2 \subset \mathbb{R}^n$ be closed and convex. Suppose $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable, and strongly convex with modulus $\sigma > 0$. Furthermore, assume $x_1^\star, x_2^\star \in X_1 \cap X_2$, where $x_i^\star = \arg\min\{f_i(x) \mid x \in X_i\}$ for $i = 1, 2$. If

$$\exists \delta_1 \geq 0,\ \forall x \in X_1 \cap X_2,\ \|\nabla f_1(x) - \nabla f_2(x)\| \leq \delta_1,$$

then $\|x_2^\star - x_1^\star\| \leq \frac{\delta_1}{\sigma}$.

Proof: The result is trivial if $x_1^\star = x_2^\star$, so assume $v := x_2^\star - x_1^\star \neq 0$. Since f_1, f_2 are convex,

$$\forall i \in \{1,2\},\ \forall x \in X_i,\ \nabla f_i(x_i^\star)^\top (x - x_i^\star) \geq 0. \tag{21}$$

Choosing $i = 2$ and $x = x_1^\star \in X_2$, we obtain

$$\nabla f_2(x_2^\star)^\top v \leq 0. \tag{22}$$

By the mean value theorem,

$$\nabla f_2(x_2^\star) = \nabla f_2(x_1^\star) + \nabla^2 f_2(x_1^\star + Tv)v,$$

for some $T = \operatorname{diag}(t_1, ..., t_n)$, where $t_1, ..., t_n \in (0,1)$. Substituting this result into (22),

$$\nabla f_2(x_1^\star)^\top v + v^\top \nabla^2 f_2(x_1^\star + Tv)v \leq 0,$$

which implies

$$v^\top \nabla^2 f_2(x_1^\star + Tv)v + \nabla f_1(x_1^\star)^\top v \leq \left[\nabla f_1(x_1^\star) - \nabla f_2(x_1^\star)\right]^\top v. \tag{23}$$

Now, choosing $i = 1$ and $x = x_2^\star \in X_1$, (21) also implies $\nabla f_1(x_1^\star)^\top v \geq 0$. Thus, by strong convexity,

$$\begin{aligned}\sigma\|v\|^2 &\leq v^\top \nabla^2 f_2(x_1^\star + Tv)v \\ &\leq v^\top \nabla^2 f_2(x_1^\star + Tv)v + \nabla f_1(x_1^\star)^\top v.\end{aligned}$$

Combining this with (23) then yields

$$\sigma\|v\|^2 \leq \left[\nabla f_1(x_1^\star) - \nabla f_2(x_1^\star)\right]^\top v \leq \delta_1\|v\|,$$

which implies the result. ■

This result is a direct generalisation of [12, (A.16)] to the constrained case. We observe that two conditions are sufficient for bounding the shift in minimisers.

Assumption 6: There exists $\delta_1 \geq 0$ for the pair $(\mathcal{F}, \mathcal{X})$, such that

$$\forall k,\ \forall x \in X_k \cap X_{k+1},\ \|\nabla f_{k+1}(x) - \nabla f_k(x)\| \leq \delta_1.$$

Assumption 7: Let $(\mathcal{F}, \mathcal{X})$ be such that $x_{k-1}^\star, x_{k+1}^\star \in X_k$.

Remark 4.1: In practice, it is difficult to guarantee $x_{k+1}^\star \in X_k$, however, this follows automatically if $X_{k+1} \subset X_k$.

Proposition 4.3: Let $(\mathcal{F}, \mathcal{X})$ satisfy Assumptions 1 – 3, 6 and 7. Then Assumption 4 is also satisfied with $V = \frac{\delta_1}{\sigma}$.

Proof: Assumption 7 is equivalent to $x_k^\star, x_{k+1}^\star \in X_k \cap X_{k+1}$. The rest follows directly from Lemma 4.2. ■

C. Sub-optimality bounds under compact constraints

One way of obtaining sub-optimality bounds in the presence of constraints is to impose a uniform Lipschitz property on the f_k over X_k. Given strong convexity, this implies the compactness of every feasible set.

Assumption 8 (Uniform Lipschitz cost): For the pair $(\mathcal{F}, \mathcal{X})$, there exists $M > 0$ such that

$$\forall k \in \mathbb{N}_0,\ \forall x \in X_k,\ \|\nabla f_k(x)\| \leq M.$$

Lemma A.2 then gives us the inequalities

$$\phi_k \leq M e_k \tag{24}$$

$$\bar{\phi}_k \leq M \bar{e}_k. \tag{25}$$

Corollary 4.4 (Steady-state sub-optimality): Let $\mathcal{F}$, $\mathcal{X}$ satisfy Assumptions 1 – 4 and 8. Let $x_0 \in \mathbb{R}^n$, and consider iterations of the form

$$x_{k+1} = P_{X_k}\left(x_k - \alpha \nabla f_k(x_k)\right), \tag{26}$$

where $\alpha = \frac{2}{\sigma+L}$. Then

$$\limsup_{k\to\infty} \phi_k \leq \frac{MV(L+\sigma)}{2\sigma}. \tag{27}$$

$$\limsup_{k\to\infty} \bar{\phi}_k \leq \frac{MV(L-\sigma)}{2\sigma}. \tag{28}$$

A different bound on predicted sub-optimality can be obtained, if we choose step sizes $\alpha_k \leq \frac{1}{L}$.

Proposition 4.5: Let $\mathcal{F}, \mathcal{X}$ satisfy Assumptions 1 – 3 and 8. Let $x_0 \in \mathbb{R}^n$, and consider iterations of the form

$$x_{k+1} = P_{X_k}\left(x_k - \alpha_k \nabla f_k(x_k)\right),$$

where $\alpha_k \in (0, \frac{1}{L}]$ for all $k \in \mathbb{N}_0$. Then for all k,

$$\bar{\phi}_k \leq M\sqrt{M^2\alpha_k^2 + e_k^2} - M^2\alpha_k. \tag{29}$$

Proof: First note that, since $\alpha_k \in (0, \frac{1}{L}]$, [1, Proposition 6.1.6] implies [1, (6.14)] is satisfied. Thus, all the hypotheses of [1, Proposition 6.1.7] are satisfied. Applying [1, (6.17)],

$$\begin{aligned}\bar{e}_k^2 &= \|x_k^\star - x_{k+1}\|^2 \\ &\leq \|x_k^\star - x_k\|^2 - 2\alpha_k(f_k(x_{k+1}) - f_k(x_k^\star)) \\ &= e_k^2 - 2\alpha_k\bar{\phi}_k.\end{aligned}$$

The inequality (25) then implies

$$\bar{\phi}_k^2 \leq M^2(e_k^2 - 2\alpha_k\bar{\phi}_k),$$

which can be rearranged to form

$$\bar{\phi}_k^2 + 2M^2\alpha_k\bar{\phi}_k - M^2e_k^2 \leq 0.$$

Taking the non-negative solutions to the above inequality yields the result. ■

Corollary 4.6: If Assumption 4 is also satisfied, and $\alpha_k = \alpha \in (0, \frac{1}{L}]$ for all k, then

$$\limsup_{k\to\infty} \bar{\phi}_k \leq M\sqrt{M^2\alpha^2 + \frac{V^2}{(1-\rho)^2}} - M^2\alpha, \tag{30}$$

where $\rho = \max\{|1-\alpha L|, |1-\alpha\sigma|\}$.

Proof: This follows from Corollary 3.4. ■

Corollary 4.7: In particular, if $\alpha = \frac{1}{L}$, then

$$\limsup_{k\to\infty} \bar{\phi}_k \leq M\sqrt{\frac{M^2}{L^2} + \frac{V^2L^2}{\sigma^2}} - \frac{M^2}{L}. \tag{31}$$

Remark 4.2: The function $\sqrt{x^2+b} - x$ is decreasing in x and increasing in b for $x, b > 0$. Furthermore, ρ is decreasing in α for $\alpha \in (0, \frac{1}{L}]$. This implies the bound in (30) is decreasing in α, and the optimal choice of step size is therefore $\alpha = \frac{1}{L}$.

Remark 4.3: The bounds in (28) and (31) are different bounds, which hold under different choices of step size. The minimum of the two depends on the parameters M, σ, δ_1, V and L.

V. Numerical Experiments

The theoretical error and sub-optimality bounds of the previous sections are illustrated here by means of two numerical examples. In particular, Section V-B presents a non-trivial example with time-varying constraints that satisfies all the required assumptions.

A. Unconstrained Example

In this example, we use a sequence of cost functions of the form

$$f_k(x) = \tfrac{1}{2}(x - s_k)^\top Q_k(x - s_k),$$

where $Q_k \in \mathbb{R}^{n\times n}$ and $s_k \in \mathbb{R}^n$ are randomly generated and satisfy

$$\sigma I \preceq Q_k = Q_k^\top \preceq LI \tag{32}$$

$$\|s_{k+1} - s_k\| \leq V, \tag{33}$$

for all k. Results for parameter values $n = 3,\ \sigma = 5,\ L = 10,\ V = 4$ are plotted in Figure 1, which compares actual performance with the bounds dictated by (9), (7), (16) and (17), for a step size $\alpha = \frac{2}{\sigma+L}$.

B. Constrained Example

Here, we use a sequence of cost functions of the form

$$f_k(x) = \tfrac{1}{2}x^\top Qx + q_k^\top x$$

where $\sigma I \preceq Q = Q^\top \preceq LI$, and q_k is randomly generated to satisfy

$$\forall k,\ \|q_{k+1} - q_0\| \leq \delta. \tag{34}$$

The sequence of constraints is of the form

$$X_0 = \{x \in \mathbb{R}^n \mid -u \leq x \leq u\} \tag{35}$$

$$\forall k \in \mathbb{N},\ X_k = \{x \in \mathbb{R}^n \mid A_kx \leq b_k,\ -u \leq x \leq u\}, \tag{36}$$

where $u \in \mathbb{R}^n$ defines the initial box constraints,

$$A_{k+1} = \begin{bmatrix} A_k \\ a_{k+1}^\top \end{bmatrix}, \quad b_{k+1} = \begin{bmatrix} b_k \\ \|a_{k+1}\|^2 \end{bmatrix}, \tag{37}$$

and $a_{k+1} \in \mathbb{R}^n$ is randomly generated based on the previous minimiser to satisfy

$$\|a_{k+1}\| \geq \|x_k^\star\| + R, \tag{38}$$

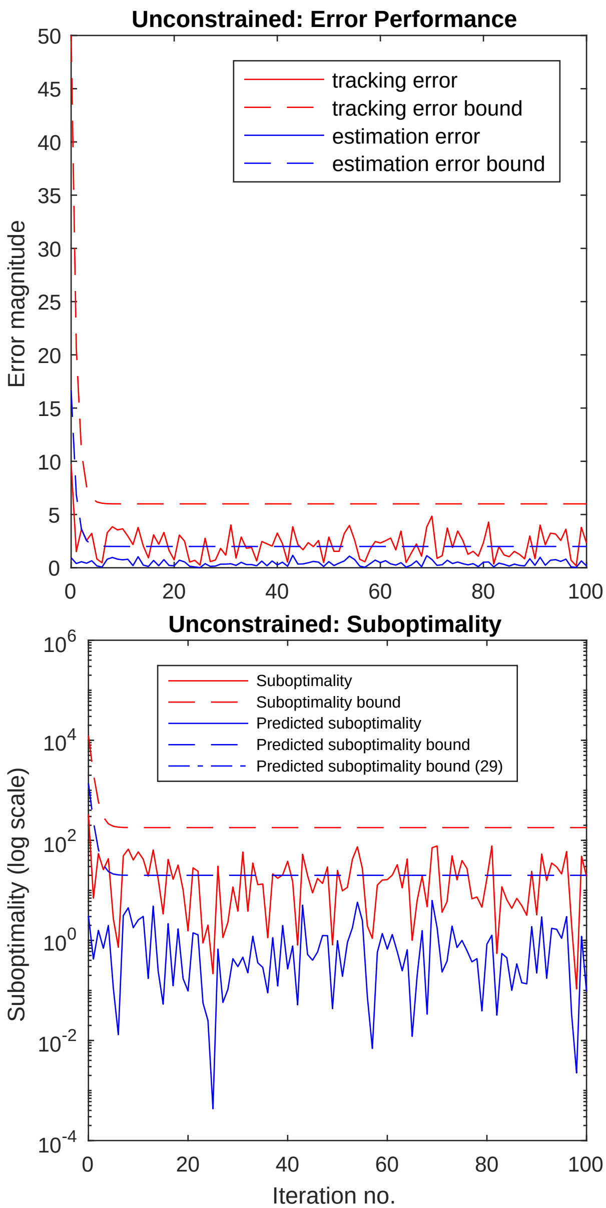

Fig. 1. Unconstrained numerical example ($\alpha = \frac{2}{\sigma+L}$)

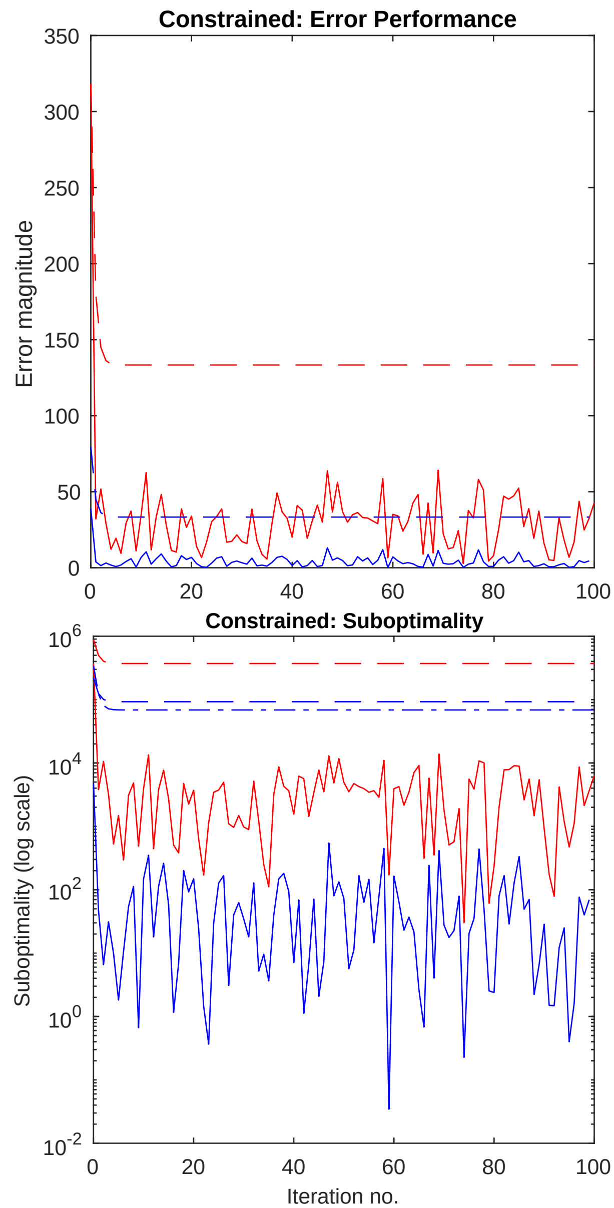

Fig. 2. Constrained numerical example ($\alpha = \frac{1}{L}$)

for some $R > \frac{2\rho\delta}{\sigma(1-\rho)}$.

Claim 5.1: Assumptions 1 – 8 hold.

Proof: Assumptions 1 – 3 are obviously true. Note that (34) implies Assumption 6 holds with constant $\delta_1 = 2\delta$. Now (37) implies $x^\star_{k+1} \in X_{k+1} \subset X_k$, and observe that

$$A_{k+1}x^\star_k = \begin{bmatrix} A_k x^\star_k \\ a^\top_{k+1} x^\star_k \end{bmatrix} \leq \begin{bmatrix} b_k \\ \|a_{k+1}\|\|x^\star_k\| \end{bmatrix} \leq \begin{bmatrix} b_k \\ \|a_{k+1}\|^2 \end{bmatrix} = b_{k+1},$$

by which $x^\star_k \in X_{k+1}$. This implies Assumption 7 is satisfied. By Proposition 4.3, Assumption 4 is then satisfied with $V = \frac{\delta_1}{\sigma}$. Let $y \in \{x \in X_k \mid \|x - x^\star_k\| \leq R\}$. Thus

$$\|y\| = \|y - x^\star_k + x^\star_k\| \leq R + \|x^\star_k\|.$$

Since $y \in X_k$,

$$A_{k+1}y = \begin{bmatrix} A_k y \\ a^\top_{k+1} y \end{bmatrix} \leq \begin{bmatrix} b_k \\ \|a_{k+1}\|(R + \|x^\star_k\|) \end{bmatrix} \leq \begin{bmatrix} b_k \\ \|a_{k+1}\|^2 \end{bmatrix} = b_{k+1},$$

by (38), and thus $y_k \in X_{k+1}$. Noting that $R > \frac{V\rho}{1-\rho}$, this establishes Assumption 5. Finally, compactness of the initial box constraints guarantees

$$\|\nabla f_k(x)\| = \|Qx_k + q_k\| \leq L\|u\| + \|q_0\| + \delta := M,$$

thereby satisfying Assumption 8. ■

Results for parameter values $n = 2,\ \sigma = 6,\ L = 8,\ \delta_1 = 300,\ u = 150 \cdot \mathbf{1} \in \mathbb{R}^n,\ R = 33.34$ are plotted in figure 2, which compares the actual performance with the bounds dictated by (9), (7), (24), (25) and (29), for a step size of $\alpha = \frac{1}{L}$. An animation of this can be found at `https://youtu.be/DV7Jb5IQDms`.

VI. Conclusion

The performance of projected gradient descent on a class of time-varying optimisation problems is considered in this paper. Attention is restricted to time-varying strongly convex functions with Lipschitz gradient, subject to time-varying, convex constraints. Tracking and estimation errors are defined, along with the analogous sub-optimality and predicted sub-optimality levels. Bounds on each of these quantities are derived. Conditions on the sequence of constraints are also presented that guarantee finite-time feasibility, and a bounded distance between successive optimisers. This theoretical analysis is then illustrated and verified using numerical examples. Future works will consider gradient-free methods for time-varying optimisation, and extensions to unconstrained nonlinear problems that relax the convexity assumptions.

References

[1] D. P. Bertsekas. *Convex Optimization Algorithms*. Athena Scientific Belmont, 2015.

[2] R. L. G. Cavalcante and S. Stanczak. A Distributed Subgradient Method for Dynamic Convex Optimization Problems Under Noisy Information Exchange. *IEEE Journal of Selected Topics in Signal Processing*, 7(2):243–256, Apr. 2013.

[3] J. W. Daniel. Stability of the solution of definite quadratic programs. *Mathematical Programming*, 5(1):41–53, Dec. 1973.

[4] A. Dontchev, M. Krastanov, R. Rockafellar, and V. Veliov. An Euler–Newton Continuation Method for Tracking Solution Trajectories of Parametric Variational Inequalities. *SIAM Journal on Control and Optimization*, 51(3):1823–1840, Jan. 2013.

[5] E. Hazan. *Introduction to Online Convex Optimization*. 2017. OCLC: 995130886.

[6] F. Y. Jakubiec and A. Ribeiro. D-MAP: Distributed Maximum a Posteriori Probability Estimation of Dynamic Systems. *IEEE Transactions on Signal Processing*, 61(2):450–466, Jan. 2013.

[7] Q. Ling and A. Ribeiro. Decentralized Dynamic Optimization Through the Alternating Direction Method of Multipliers. *IEEE Transactions on Signal Processing*, 62(5):1185–1197, Mar. 2014.

[8] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1997.

[9] J. H. Manton. A framework for generalising the Newton method and other iterative methods from Euclidean space to manifolds. *Numerische Mathematik*, 129(1):91–125, Jan. 2015.

[10] A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro. Online Optimization in Dynamic Environments: Improved Regret Rates for Strongly Convex Problems. *arXiv:1603.04954 [cs, math]*, Mar. 2016.

[11] M. J. Neely and H. Yu. Online Convex Optimization with Time-Varying Constraints. *arXiv:1702.04783 [math]*, Feb. 2017.

[12] A. Y. Popkov. Gradient Methods for Nonstationary Unconstrained Optimization Problems. *Automation and Remote Control*, 66(6):883–891, June 2005.

[13] A. Simonetto. Time-Varying Convex Optimization via Time-Varying Averaged Operators. *arXiv:1704.07338 [math]*, Apr. 2017.

[14] A. Simonetto and G. Leus. Distributed asynchronous time-varying constrained optimization. In *2014 48th Asilomar Conference on Signals, Systems and Computers*, pages 2142–2146, Nov. 2014.

[15] A. Simonetto and G. Leus. Double smoothing for time-varying distributed multiuser optimization. In *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 852–856, Dec. 2014.

[16] A. Simonetto, A. Mokhtari, A. Koppel, G. Leus, and A. Ribeiro. A Class of Prediction-Correction Methods for Time-Varying Convex Optimization. *IEEE Transactions on Signal Processing*, 64(17):4576–4591, Sept. 2016.

[17] C. Sun, M. Ye, and G. Hu. Distributed Time-Varying Quadratic Optimization for Multiple Agents Under Undirected Graphs. *IEEE Transactions on Automatic Control*, 62(7):3687–3694, July 2017.

[18] S. Y. Tu and A. H. Sayed. Mobile Adaptive Networks. *IEEE Journal of Selected Topics in Signal Processing*, 5(4):649–664, Aug. 2011.

[19] C. Xi and U. A. Khan. Distributed dynamic optimization over directed graphs. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 245–250, Dec. 2016.

[20] M. Ye and G. Hu. Distributed optimization for systems with time-varying quadratic objective functions. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 3285–3290, Dec. 2015.

[21] V. Zavala and M. Anitescu. Real-Time Nonlinear Optimization as a Generalized Equation. *SIAM Journal on Control and Optimization*, 48(8):5444–5467, Jan. 2010.

[22] M. M. Zavlanos, A. Ribeiro, and G. J. Pappas. Network Integrity in Mobile Robotic Networks. *IEEE Transactions on Automatic Control*, 58(1):3–18, Jan. 2013.

[23] M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 928–936, 2003.

Appendix

A. Technical Lemmata

Lemma A.1: If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and satisfies

$$\exists L > 0,\ \forall x, y \in \mathbb{R}^n,\ \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|,$$

then for all $x, x^\star \in \mathbb{R}^n$,

$$f(x) - f(x^\star) \leq \|\nabla f(x^\star)\|\|x - x^\star\| + \tfrac{1}{2}L\|x - x^\star\|^2.$$

Proof: Let $e := x - x^\star$. Then by Taylor's theorem, there exists $t \in (0, 1)$ such that

$$f(x) = f(x^\star) + \nabla f(x^\star)^\top e + \frac{1}{2}e^\top \nabla^2 f(x^\star + te)e.$$

The result then follows. ■

Lemma A.2: Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Then

$$\forall x, x^\star \in \mathbb{R}^n,\ f(x) - f(x^\star) \leq M\|x - x^\star\|,$$

where $M := \sup\{\|\nabla f(x)\| \mid x \in X\}$.

Proof: By the mean value theorem, there exists $t \in (0, 1)$ such that

$$f(x) = f(x^\star) + \nabla f\big(tx + (1 - t)x^\star\big)^\top (x - x^\star).$$

■