Time-Varying Multi-Objective Optimization: Tradeoff Regret Bounds

Allahkaram Shafiei[1[0000−0002−1415−2166]], Vyacheslav Kungurtsev[1[0000−0003−2229−8824]], and Jakub Marecek[1[0000−0003−0839−0691]]

Czech Technical University Prague
shafiall@fel.cvut.cz
kunguvya@fel.cvut.cz
jakub.marecek@fel.cvut.cz

Abstract. Multi-objective optimization studies the process of seeking multiple competing desiderata in some operation. Solution techniques highlight marginal tradeoffs associated with weighing one objective over others. In this paper, we consider time-varying multi-objective optimization, in which the objectives are parametrized by a continuously varying parameter and a prescribed computational budget is available at each time instant to algorithmically adjust the decision variables to accommodate for the changes. We prove regret bounds indicating the relative guarantees on performance for the competing objectives.

Keywords: Time-varying · Multi-objective · Proximal Gradient Decent· Regret Bound.

1 Introduction and Preliminaries

During the last decades, substantial research efforts have been devoted to learning and decision-making in environments with functionally relevant streaming data with potentially changing statistical properties. In many engineering design problems with social impact, including optimal power flow and sensor networks [11, 14], mobile robots [18], and non-linear distributed flow equations [12], there are potentially multiple criteria to consider in characterizing the best learning or decision-making performance. Formally, such multi-criteria optimization problems are classed as multi-objective optimization.

The setting of a dynamically changing and uncertain environment lends itself to what is classed as online optimization, where the cost function changes over time and an adaptive decision pertaining only to past information has to be made at each stage. The standard convergence criteria in online optimization is the level of *regret*, a quantity capturing the difference between the accumulated cost incurred up to some arbitrary time and the cost obtained from the best fixed point chosen in hindsight.

In machine-learning applications of multi objective optimization, the time-varying aspects could capture, e.g., time-varying group structure, seasonal or circadian cyclicity, or some form of a concept drift. In game theory, the time-varying aspects could capture time-varying pay-offs (or time-varying price elasticity of the demand) in extensive forms of Stackelberg-like games or time-varying demands in congestion games.

Our contributions include:

- introduction of regret tradeoffs as the appropriate metric for grading solvers for online multi-objective optimization
- an on-line proximal-gradient algorithm for handling multiple time-varying convex objectives,
- theoretical guarantees for the algorithm.

2 Related work

Proximal Gradient Descent is a natural approach for minimizing both single and multiple objectives. One of the most widely studied methods for multiobjective optimization problems is steepest descent, e.g., [1, 7].

Subsequently, a proximal point method [8], that can be applied to non-smooth problems, was considered. However, this method is just a conceptual scheme and does not necessarily generate subproblems that are easy to solve. For non-smooth problems, a subgradient method was also developed [10]. A very comprehensive recent paper [19] has presented the regret bounds for classic algorithms for online convex optimization with Lipschitz, but possibly non-differentiable functions, proving a regret of $O(\frac{1}{\sqrt{K}})$, with K iterations at each time instant. With respect to multiobjective (but not online) optimization, H.Tanabe [9] proposed proximal gradient methods with and without line searches for unconstrained multiobjective optimization problems, in which every objective function is of the composite form of interest in our work, $F_i(x) = f_i(x) + g_i(x)$, with f_i smooth and g_i merely proper and convex but with a tractable proximal computation.

Next, we describe the literature on online time-varying convex single objective optimization. As the first innovative paper in this space, Zinkevich [15] proposed a gradient descent algorithm with a regret bound of $O(\sqrt{K})$. In the case that cost functions are strongly convex, the regret bound of the online gradient descent algorithm was further reduced to be $O(\log(K))$ with suitably chosen step size by several online algorithms presented in [5].

3 Problem Formulation

We begin with describing the problem of Time-Varying Multi-Objective Optimization. Suppose we have a sequence of convex cost functions $\phi_{i,t}(x) := f_{i,t}(x) + g_{i,t}(x)$ where $f_{i,t}$ are smooth and $g_{i,t}$ non-smooth. The index t corresponds to the time step, and i indexes the objective function among the set of desiderata. ABetween each time step $t \in [T] := \{1, 2, ..., T\}$, there is a finite amount of time available to compute an optimal decision, indicated by a maximum iteration count for any Algorithm, at which point the decision maker must choose an action $x^t \in R$, and the decision maker is faced with a loss of $\phi_{i,t}(x^t)$. In this scenario, due to insufficient computation time, the decision does not necessarily correspond to the minimizers, and the decision maker faces a so-called regret. Regret is defined as the difference between the accumulated cost over time and the cost incurred by the best-fixed decision when all functions are known in advance, see [5, 6, 15]. Let us consider $\phi_{i,t}(x) = f_{i,t}(x) + g_{i,t}(x)$ as

$$F(x,t) = (\phi_{1,t}(x), \phi_{2,t}(x), ..., \phi_{N,t}(x))$$

At the time t, we consider the following time-varying vector optimization

$$\min_{x \in R^n} \ (F(x,t) := F_t(x)) \tag{1}$$

where $F : R^n \to R^m$ and, each $f_{i,t}$ is $L_{f_{i,t}}$ Lipschitz continuously differentiable and g_i is convex with a simple prox evaluation. Throughout the entire paper, our discussion on (1) and the proposed algorithm are motivated by the following: As can be surmised from the definition, there is rarely a singleton that is a Pareto optimal point. Usually, there is a continuum of solutions. As such one can consider a *Pareto front* which indicates the set of Pareto optimal points. The front represents the objective values reached by the components of the range of $F(x)$ and it is usually a surface of $m-1$ dimensions. One can consider it as representing the tradeoffs associated with the optimization problem, to lower i's value, i.e. $f_i(x)$, how much are you willing to compromise in terms of potentially raising $f_j(x)$ for the set of $j \neq i$? Because of the fundamental generality of the concept of a solution to a vector optimization problem, finding a solution can be defined as (see, e.g., for a survey [17] and for a text [13]):

1. Visualizing the entire Pareto front, or some portion of it
2. Finding *any* point on the Pareto front
3. Finding some point that satisfies an additional criteria, effectively making this a bilevel optimization problem.

In regards to the second option, one can notice that this can be done, in the convex case, by solving the so called "scalarized" problem:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^N \omega_i f_i(x), \ \sum_{i=1}^N \omega_i = 1, \ 0 \le \omega_i \le 1$$

for any valid choice of $\{\omega_i\}$. This reduces the problem to simple unconstrained optimization. This leaves the choice of said constants, however, arbitrary, and thus not all that informative. Although the parameters are weights balancing the relative importance of the objective functions, poor relative scaling across $f_i(x)$ can make an informed choice of $\{\omega_i\}$ insurmountable. For example, if $f_1(x) = 1000x^2$ and $f_2(x) = 0.001(x-2)^2$, taking $\omega_1 = \omega_2 = 0.5$ clearly pushes the solution of the scalarized problem to prioritize minimizing $f_1(x)$.

As an additional challenge, we consider the time varying case, i.e., each $f_i(x)$ changes over time, e.g., due to data streaming with concept drift. With a finite processing capacity at each time instant, we seek an Algorithm that appropriately balances the objectives at each time instant.

In this paper, we introduce scalarization at the *algorithmic* level for time varying multiobjective optimization. In particular, at each iteration, we consider computing a set of steps, each of which intends to push an iterate towards the solution of the problem of minimizing $f_i(x)$ exclusively. The algorithm then form a convex combination of these steps with a priori chosen coefficients. We derive *tradeoff regret bounds* indicating how the choice of said coefficients results in guarantees in regards to suboptimality for every objective. We assert that this would be the most transparently informative theoretical guarantee, in terms of exactly mapping algorithmic choices to comparative performance for every objective function, and as such a natural and important contribution to time varying multi-objective optimization.

Now we shall present our formal assumptions in regards to the problem, in particular the functional properties of F as well as the Algorithm we are proposing and studying the properties of in this paper.

Assumption 1 (Problem Structure) (i) *For all i, t functions $f_{i,t}(\cdot) : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable such that the gradient is Lipschitz with constant $L_{f_{i,t}}$:*

$$\|\nabla f_{i,t}(x) - \nabla f_{i,t}(y)\| \leq L_{f_{i,t}} \|x - y\|, \quad \forall x, y \in \mathbb{R}^n$$

(ii) *For all t, the function $g_{i,t}(\cdot) : \mathbb{R}^n \to (-\infty, \infty]$ is proper, lower semi-continuous, and convex, but not necessarily differentiable. Also, assume that $\mathrm{dom}(g_{i,t}(\cdot)) = \{x \in \mathbb{R}^n : \ g_{i,t}(x) < \infty\}$ is non-empty and closed.*

(iii) *Corresponding to each objective $\phi_{i,t}$ we consider $T_{i,t}(x) = \mathrm{prox}_{C_i g_{i,t}}(x - C_i \nabla f_{i,t}(x))$*

We also assume a bound on the magnitude of change between successive times:

Assumption 2 (Slow Changes) *The observations as compared to estimates of the function values from the previous time step are bounded at all x, i.e.,*

$$\sup_{t \geq 1} \max_{i \in [N]} \left\{ \begin{array}{l} |f_{i,t+1}(x) - f_{i,t}(x)|, \\ |g_{i,t+1}(x) - g_{i,t}(x)| \end{array} \right\} \leq e$$

4 Algorithm and Preliminaries

The Algorithm is stated formally as Algorithm 1. The coefficients $\{\alpha_i\}$ denote the priority of objective i, and belong to the unit simplex (meaning $\sum_{i=1}^{N} \alpha_i = 1$, $0 \leq \alpha_i \leq 1$).

The following assumptions are typical in the analysis of online algorithms and make real-time algorithmic path-following of solutions feasible. In particular, we consider the online streaming setting with a finite sampling rate, which we assume permits K iterations of the proximal-gradient steps between two updates of the inputs.

Algorithm 1 On-Line Multi-Objective Proximal Gradient Decent

Input: Initial iterate x^1 solving the problem with data $f_{1,1}(x), g_{1,1}(x)$ parameters $C_1 \in (0, \frac{1}{L_{f_{1,1}}}]$, $\alpha_i > 0$, and let $x^{1,0} \leftarrow x^1$

for $t = 1, 2, ..., T$ **do**

$x^{t,1} \leftarrow x^t$;

Receive data $f_{i,t}(x^t), g_{i,t}(x^t)$;

for $k = 0, 1, 2, ..., K$ **do**

$y^{t,k+1,i} \leftarrow \text{prox}_{C_i g_{i,t}}(x^{t,k} - C_i \nabla f_{i,t}(x^{t,k}))\, \forall i$;

$x^{t,k+1} \leftarrow \sum_{i=1}^{N} \alpha_i y^{t,k+1,i}$;

$k \leftarrow k+1,\ t \leftarrow t+1$

end for

$x^{t+1,0} \leftarrow x^{t,K}$ and $x^{t+1} \leftarrow x^{t,K+1}$;

end for

Assumption 3 (Sufficient Processing Power) *At all times $t \in [T]$, the algorithm executes at least K iterations before receiving the new input.*

We consider two measures of the quality of the solution trajectory:

(A): The *dynamic regret bound* (see, e.g., [15], and reference therein) defined as:

In the case of static regret [4], $x^{\text{opt},t,i}$ is replaced by $x^{\text{opt},i} \in \text{argmin}_{x \in X} \sum_{t=1}^{T} \phi_{i,t}(x)$, i.e,

$$\text{S} - \text{Reg}_i = \sum_{t=1}^{T} \phi_{i,t}(x^t) - \min_{x \in X} \sum_{t=1}^{T} \phi_{i,t}(x)$$

(B): In addition, we will consider the following quantities:

$$\begin{aligned} \phi_t(x) &:= \sum_{i \in [N]} \alpha_i \phi_{i,t}(x), \quad x^{\text{opt},t} \in \operatorname*{argmin}_x \phi_t(x) \\ W_T &:= \sum_{t \in [T]} \|x^{\text{opt,t+1}} - x^{\text{opt,t}}\|^2 \end{aligned}$$

Now we define the dynamic of the regret bound for the convex combination of $\phi_{i,t}$ as follows

$$\text{Reg}_t = \sum_{t=1}^{T} \phi_t(x^t) - \sum_{t=1}^{T} \phi_t(x^{\text{opt},t})$$

Assumption 4 *Choose α_i such that $T_t(x) = \sum_{i \in [N]} \alpha_i T_{i,t}(x)$ is proximal operator of $\phi_t(x)$.*

The following lemma is a key result throughout the paper.

Lemma 1. *Let f be convex and smooth, and g be non-smooth and $\phi = f + g$ then*

$$\phi(T(x)) - \phi(y) \leq \frac{1}{2C}[\|x-y\|^2 - \|T(x)-y\|^2] \tag{2}$$
$$\text{and} \quad \phi(T(x)) \leq \phi(x). \tag{3}$$

where $T(x) = \mathrm{prox}_{Cg}(x - C\nabla f(x))$, $C \in (0, \frac{1}{L_f}]$ and L_f is Lipschitz constant for ∇f.

Proof. Take $G(x) := \frac{1}{C}(x - T(x))$ and apply the standard Descent Lemma:

$$f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L_f}{2}\|x-y\|^2 \quad \forall x, y \in \mathbb{R}^n. \tag{4}$$

Plugging $y = x - CG(x)$ in (4) one obtains that

$$f(x - CG(x)) \leq f(x) + \nabla f(x)^T((x - CG(x)) - x) + \frac{L_f C^2}{2}\|G(x)\|^2 \tag{5}$$
$$\leq f(x) - C\nabla f(x)^T(G(x)) + \frac{C}{2}\|G(x)\|^2. \tag{6}$$

Now, from $x - CG(x) = \mathrm{prox}_{Cg}(x - C\nabla f(x))$ it follows that

$$G(x) - \nabla f(x) \in \partial g(x - CG(x)).$$

Therefore, for any y, by convexity of g we obtain the relation:

$$g(x - CG(x)) \leq g(y) - (G(x) - \nabla f(x))^T(y - x - CG(x)). \tag{7}$$

Now consider $\phi(T(x)) = \phi(x - CG(x)$. By simplifying and applying (7), one has

$$\begin{aligned}
\phi(x - CG(x)) &= f(x - CG(x)) + g(x - CG(x)) \\
&\leq f(x) - C\nabla f(x)^T(G(x)) + \frac{C}{2}\|G(x)\|^2 + g(x - CG(x)) \\
&\leq f(y) - \nabla f(x)^T(y-x) - C\nabla f(x)^T(G(x)) + \frac{C}{2}\|G(x)\|^2 + g(x - CG(x)) \\
&\leq f(y) - \nabla f(x)^T(y-x) - C\nabla f(x)^T(G(x)) \\
&+ \frac{C}{2}\|G(x)\|^2 + g(y) - (G(x) - \nabla f(x))^T(y - x + CG(x)) \\
&= \phi(y) - \nabla f(x)^T(y-x) - C\nabla f(x)^T(G(x)) + \frac{C}{2}\|G(x)\|^2 - G(x)^T(y-x) \\
&- C\|G(x)\|\nabla f(x)^T(y-x) + C\nabla f(x)^T(G(x)) \\
&\leq \phi(y) + \frac{1}{2C}[\|x-y\|^2 - \|(x-y) - CG(x)\|^2],
\end{aligned}$$

5 Main Results

Our main result provides a bound on the expected dynamic regret of the online multi-objective proximal gradient descent (Algorithm 1). Depending on the coefficients α_i, there are two cases, i) if for all $i \in [N]$, $\alpha_i \neq 0$ and ii) if there is $i \in [N]$ such that $\alpha_i = 1$ and for all $j \neq i$, $\alpha_j = 0$. For case i), we have

Theorem 1. *Let x^t, (t=1,..., T) be a sequence generated by running Algorithm 1 over T time steps. Then, we have*

$$Reg = \sum_{t=1}^{T} \phi_t(x^t) - \sum_{t=1}^{T} \phi_t(x^{\mathrm{opt},t}) \leq C_T + 4(T-1)e + \|x^1 - x^{\mathrm{opt},1}\|^2 + W_T$$

where $C_T = |\phi_1(x^1) - \phi_1(x^{\mathrm{opt},\mathrm{T}})|$. In addition one has,

$$Reg = \sum_{t=1}^{T} \phi_{i,t}(x^t) - \sum_{t=1}^{T} \phi_{i,t}(x^{\mathrm{opt},t}) \leq \frac{1}{\alpha_i}\left[C_T + 4(T-1)e + \|x^1 - x^{\mathrm{opt},1}\|^2 + W_T\right]$$

To prove the result, we need a technical lemma:

Lemma 2. *The following holds.*
(a) For all $t \in [T], k \in [K]$ one has

$$\|x^{t,k+1} - x^{\mathrm{opt,t}}\| \leq \|x^{t,k} - x^{\mathrm{opt,t}}\|$$

(b) For all $t \in [T]$ one has $\phi_t(x^{t+1}) \leq \phi_t(x^t)$ and particularly $|\phi_t(x^{t+1}) - \phi_{t+1}(x^t)| < e$
(c) For all $t \in [T]$, one has $|\phi_t(x) - \phi_{t+1}(x)| < 2e$

Returning to the proof of the main result,

Proof. Utilizing Lemma 1, one obtains

$$\begin{aligned}
&\phi_t(x^{t+1}) - \phi_t(x^{\mathrm{opt,t}}) && (8)\\
&= \phi_t(T_t(x^{t,K})) - \phi_t(x^{\mathrm{opt,t}}) \leq \frac{1}{\widetilde{C}}\left[\|x^{t,K} - x^{\mathrm{opt,t}}\|^2 - \|T_t(x^{t,K}) - x^{\mathrm{opt,t}}\|^2\right]\\
&= \frac{1}{\widetilde{C}}\left[\|x^{t,K} - x^{\mathrm{opt,t}}\|^2 - \|x^{t+1} - x^{\mathrm{opt,t}}\|^2\right] \leq \frac{1}{\widetilde{C}}\left[\|x^{t,1} - x^{\mathrm{opt,t}}\|^2 - \|x^{t+1} - x^{\mathrm{opt,t}}\|^2\right]\\
&= \frac{1}{\widetilde{C}}\left[\|x^{t} - x^{\mathrm{opt,t}}\|^2 - \|x^{t+1} - x^{\mathrm{opt,t}}\|^2\right] && (9)
\end{aligned}$$

where $T_t(x) = \sum_{i\in[N]} \alpha_i T_{i,t}(x)$, $\widetilde{C} = \sum_{i\in[N]} \alpha_i C_i$. Alternatively, it is straightforward to verify that

$$\|x^{t+1} - x^{\mathrm{opt,t}}\|^2 \geq \|x^{t+1} - x^{\mathrm{opt,t+1}}\|^2 - \|x^{\mathrm{opt,t+1}} - x^{\mathrm{opt,t}}\|^2 \quad (10)$$

the above combined with (9) leads to the following

$$\phi_t(x^{t+1}) - \phi_t(x^{\mathrm{opt,t}}) \leq \frac{1}{\widetilde{C}}\left[\|x^{t} - x^{\mathrm{opt,t}}\|^2 - \|x^{t+1} - x^{\mathrm{opt,t+1}}\|^2 + \|x^{\mathrm{opt,t+1}} - x^{\mathrm{opt,t}}\|^2\right].$$

Now, summing up the result over $t \in [T]$ derives

$$\sum_{t=1}^{T} \phi_t(x^{t+1}) - \sum_{t=1}^{T} \phi_t(x^{\text{opt,t}}) \leq \frac{1}{\widetilde{C}} \left[\|x^1 - x^{\text{opt},1}\|^2 + W_T\right] \tag{11}$$

which resulted in

$$\sum_{t=1}^{T-1} \phi_t(x^{t+1}) - \sum_{t=1}^{T-1} \phi_t(x^{\text{opt,t}}) \leq \frac{1}{\widetilde{C}} \left[\|x^1 - x^{\text{opt},1}\|^2 + W_T\right] \tag{12}$$

on the other side since $\phi_{t+1}(x^{t+1}) - 2e \leq \phi_t(x^{t+1})$ we would have the following

$$\left(\sum_{t=1}^{T} \phi_t(x^{t}) - \sum_{t=1}^{T} \phi_t(x^{\text{opt,t}})\right) - \phi_1(x^1) - 2(T-1)e + \phi_T(x^{\text{opt,T}}) \tag{13}$$

$$\leq \frac{1}{\widetilde{C}} \left[\|x^1 - x^{\text{opt},1}\|^2 + W_T\right] \tag{14}$$

taking summation over $t \in [T]$ follows that

$$\sum_{t=1}^{T} \phi_t(x^{t}) - \sum_{t=1}^{T} \phi_t(x^{\text{opt,t}}) \leq C_T + 4(T-1)e + \frac{1}{\widetilde{C}} \left[\|x^1 - x^{\text{opt},1}\|^2 + W_T\right] \tag{15}$$

In the following corollary, it will be shown that for a single objective, the dynamic regret bound is weaker than for a multi-objective case.

Corollary 1. *In the case that there exists i, $\alpha_i = 1$ and for all $j \neq i$ we have $\alpha_j = 0$ the problem reduces to time-varying single objective optimization, i.e,*

$$x^{t,k+1} = \text{prox}_{C_i g_{i,j}}(x^{t,k} - C_i \nabla f_{i,j}(x^{t,k})).$$

Then

$$\sum_{t=1}^{T} \phi_{i,t}(x^{t}) - \phi_{i,t}(x^{\text{opt},t,i}) \leq C_T + 4(T-1)e + \frac{1}{(K+1)C_i} \left[\|x^1 - x^{\text{opt},1}\|^2 + W_T\right] \tag{16}$$

Proof. As can be seen from Lemma 1 we have

$$\phi_{i,t}(x^{t,k+1}) - \phi_{i,t}(x^{\text{opt},t,i}) \leq \frac{1}{C_i}[\|x^{t,k} - x^{\text{opt},t,i}\|^2 - \|x^{t,k+1} - x^{\text{opt},t,i}\|^2].$$

Summing the result over k from 1 to K we conclude that

$$\sum_{k=1}^{K}[\phi_{i,t}(x^{t,k+1}) - \phi_{i,t}(x^{\text{opt},t,i})] \leq \frac{1}{C_i}[\|x^{t,1} - x^{\text{opt},t,i}\|^2 - \|x^{t,K+1} - x^{\text{opt},t,i}\|^2],$$

Since $\phi_{i,t}(x^{t,k+1}) \leq \phi_{i,t}(x^{t,k})$ the previous term gives that

$$\phi_{i,t}(x^{t,K+1}) - \phi_{i,t}(x^{\text{opt},t,i}) \leq \frac{1}{(K+1)C_i}[\|x^t - x^{\text{opt},t,i}\|^2 - \|x^{t+1} - x^{\text{opt},t,i}\|^2],$$

Now, by using (10) (11), and subsequently summing the previous inequality over t from 1 to T, one establishes the required assertions.

If we make an additional assumption regarding the correspondence of the function values and the distance to the solution set, we can obtain guarantees for the latter.

Assumption 5 *For all i and t, $\phi_{i,t}$ satisfies the quadratic growth property, i.e.,*

$$\phi_{i,t}(x) \geq \phi_{i,t}(x^{\text{opt},t,i}) + \frac{\gamma_{i,t}}{2}\text{dist}^2(x, S^{i,t}) \quad \text{ for all } x \in [\phi_{i,t} \leq \phi^*_{i,t} + \nu_{i,t}]$$

in which $S^{i,t}$ is the set of all optimal points of $\phi_{i,t}$, and $\phi^*_{i,t} = \phi_{i,t}(x^{\text{opt},t,i})$

The following regret bound can be readily deduced from [20, Corrolary 3.6] It is worth noting that the complexity bound aligns with the linear rate of convergence exhibited by the proximal gradient method when employed for strongly convex functions, albeit with a constant factor.

Corollary 2. *The following regret bound holds*

$$Reg_i = \sum_{t=1}^{T} \phi_{i,t}(x^t) - \sum_{t=1}^{T} \phi_{i,t}(x^{\text{opt},t,i}) \leq T\epsilon \tag{17}$$

after at most

$$t \leq \frac{\beta_T}{2\nu_T}\Gamma_T + 12\frac{\beta_T}{\gamma^0} ln\frac{M_T}{\epsilon} \quad \text{ iterations.} \tag{18}$$

where

$$M_T = \min_{t\in[T]}\min_{i\in[N]} \phi_{i,t}(x^1) - \phi_{i,t}(x^{\text{opt},t,i}),\ \Gamma_T = \min_{i\in[N]}\min_{t\in[T]} \text{dist}(x^1, S^{i,t}) \tag{19}$$

$$\beta_T = \min_{i\in[N]}\min_{t\in[T]} L_{f_{i,t}},\ \nu_T = \min_{i\in[N]}\min_{t\in[T]},\ \gamma^0 = \min_{i\in[N]}\min_{t\in[T]} \gamma_{i,t},\ \nu_T = \min_{i\in[N]}\min_{t\in[T]} \nu_{i,t}$$

Proof. First, we note that by considering Theorem 3.2 and Corollary 3.6 of [20] one can see for all i, t that

$$\phi_{i,t}(x) - \phi_{i,t}(x^{\text{opt},t,i}) \leq \epsilon \quad \forall x \in [\phi_{i,t} \leq \phi^*_{i,t} + \nu_{i,t}] \tag{20}$$

for

$$t \leq \frac{L_{f_{i,t}}}{\nu_{i,t}}\text{dist}(x^1, S^{i,t}) + 12\frac{L_{f_{i,t}}}{\gamma_{i,t}} Ln(\frac{\phi_{i,t}(x^1) - \phi_{i,t}(x^{\text{opt},t,i})}{\epsilon})$$

now, assume that x^t generated by Algorithm 1 and taking into account quantities (19), and also taking summation from $t = 1$ to $t = T$ from (20) one observes that $Reg_i \leq T\epsilon$ for at most t defined in (18).

6 Conclusions

We have studied a time-varying multi-objective optimization problem in a setting, which has not been considered previously. We have shown the properties of a natural, online proximal-gradient algorithm when the processing power between iterations is bounded. Going forward, one could clearly consider alternative uses of the same algorithm (e.g., how many operations one requires per update to achieve a certain bound in terms of the dynamic regret), variant algorithms, or completely novel settings. In parallel with our work, Tarzanagh and Balzano [3] studied online bilevel optimization under assumptions of strong convexity throughout, which could be seen as one such novel setting.

Bibliography

[1] Arthur, G.: Proper efficiency and the theory of vector maximization. J. Math. Anal and Appl. 22(3), 618–630 (1968)

[2] Allahkaram, Sh., Vyacheslav, K., Jakub, J.: Trilevel and Multilevel Optimization using Monotone Operator Theory. arXiv preprint arXiv:2105.09407, (2021)

[3] Davoud Ataee, T., Laura, B.: Online Bilevel Optimization: Regret Analysis of Online Alternating Gradient Methods. https://arxiv.org/abs/2207.02829, (2021)

[4] Hazan, E., Singh, K., Zhang, C.: Online Learning of Linear Dynamical Systems. Adv. Neural Info. Proces.Syst. 6686–6696 (2017)

[5] Hazan, E., Agarwal, A., Kale, S.: Logarithmic regret algorithms for online convex optimization. Machine Learning. 29(2), 169–192 (2007)

[6] Hazan, E., Agarwal, A., Kale, S.: Extracting certainty from uncertainty: regret bounded by variation in costs. Mach. Lear. Res (2010)

[7] Goncalves, M., Lima, F., Prudente, L.: Globally convergent Newton-type methods for multiobjective optimization. Computational Optimization and Applications. 83, 403—434 (2022)

[8] Henri, B., Alfredo, I., Svaiter, B.: Proximal methods in vector optimization. SIAM J. Optim. 15(4), 953–970 (2005)

[9] Hiroki, T., Ellen, F., Yamashita, N.: Proximal gradient methods for multiobjective optimization and their applications. Compu. Optim.Appl. 72(2), 339–361 (2019)

[10] Bagirov, A., Adil, M., Jin, L., Karmitsa.: Subgradient method for nonconvex nonsmooth optimization. J.Optim. Appl (157), 416–435 (2013)

[11] Derenick, J., John, R.: Convex optimization strategies for coordinating large-scale robot formations. IEEE 23(6), 1252–1259 (2007)

[12] Jonathan, F.B.: Optimal consensus algorithm integrated with obstacle avoidance. Math. Program. 40(1-3), 15–27 (1988)

[13] Johannes, J.: Vector optimization. Springer. (2009)

[14] Mesut, B., Felix, W.: Optimal capacitor placement on radial distribution systems. IEEE. 4(1), 725–734 (1989)

[15] Martin, Z.: Online convex programming and generalized infinitesimal gradient ascent. Proceedings of the 20th international conference on machine learning. ICML, 928–936 (2003)

[16] Nesterov, Y.: Lectures on convex optimization. Springer. (2018)

[17] Timothy, R., Arora, J.: Survey of multi-objective optimization methods for engineering. Structural and multidisciplinary optimization. 26(6),369–395 (2004)

[18] Wang, J., Xin, M.: Optimal consensus algorithm integrated with obstacle avoidance. International Journal of Systems Science. 44(1), 166–177 (2013)

[19] Yihan, Z., Portella, S., Schmidt, M., Nicholas, H.: Regret bounds without Lipschitz continuity: online learning with relative-Lipschitz losses. Advances in Neural Information Processing Systems. 33, 15823–15833 (2020)

[20] Drusvyatskiy, D., S. Lewis, A.: Error bounds, quadratic growth, and linear convergence of proximal methods. Mathematics of Operations Research. 43, 919–948 (2018)
[21] LNCS Homepage, http://www.springer.com/lncs, last accessed 2023/10/25