

Accelerated Algorithms for a Class of Optimization Problems with Constraints

Anjali Parashar Priyank Srivastava Anuradha M. Annaswamy

Abstract—This paper presents a framework to solve constrained optimization problems in an accelerated manner based on High-Order Tuners (HT). Our approach is based on reformulating the original constrained problem as the unconstrained optimization of a loss function. We start with convex optimization problems and identify the conditions under which the loss function is convex. Building on the insight that the loss function could be convex even if the original optimization problem is not, we extend our approach to a class of nonconvex optimization problems. The use of a HT together with this approach enables us to achieve a convergence rate better than state-of-the-art gradient-based methods. Moreover, for equality-constrained optimization problems, the proposed method ensures that the state remains feasible throughout the evolution, regardless of the convexity of the original problem.

I. INTRODUCTION

Several problems in a wide range of fields such as adaptive control, machine learning, and optimization [1]–[4] utilize gradient-descent based approaches. In adaptive control, decision making in the form of reducing a performance error as well as ensuring that the learning error in parameters reduces to zero as well as learning the unknown parameters of the system, both in realtime are the underlying goals. A gradient-descent approach is often employed to realize both goals, first to obtain a fast convergence of the performance error and then, to reduce the learning error. In machine learning, fast and correct training of models such as neural networks is sought after, which necessitates the reduction of an underlying loss function using a gradient-based approach. Optimization approaches require the solution of an augmented Lagrangian in an expedient manner, through a gradient-descent method. All of these solutions have to be achieved often in the presence of various constraints. Conservation equations invariably introduces equality constraints in the form of momentum, mass, or energy balance, while capacity limits introduce inequality constraints. Given the importance of the fast convergence in all these problems, there is a need for algorithms that can lead to an order of magnitude improvement in the speed of convergence, both performance and learning errors, while retaining stability. This paper proposes such an algorithm.

Recently, a class of High-order Tuners (HT) was proposed in continuous-time [5] and in discrete-time [6] for a large

class of dynamic systems for the purpose of estimation of unknown parameters. The estimation problem for this class of systems can be reformulated as a linear regression problem, where the underlying regressors correspond to various system variables that can be measured, including inputs, outputs, and states. Each of these high-order tuners was shown to result in a stable performance error when the regressors were time-varying. One of these tuners was extended in [7] for a class of nonlinear problems where the underlying error model is still based on linear regression. The motivation for these methods come from adaptive control perspectives in [8], [9] and optimization perspectives [4], [10]. In [11], these results were extended to convex functions, and shown that the high-order tuners are capable of leading to stable performance with a potential for accelerated convergence of the performance error to zero. However, all the aforementioned works focus on optimization problems without constraints. The work [12] proposes a framework to employ Nesterov’s accelerated algorithm for equality-constrained convex optimization problems.

In this paper, we extend the results of HT in [6] and [11] for optimization in the presence of both equality and inequality constraints. For equality-constrained optimization problems, we show both for convex as well as nonconvex settings, that HT-based algorithms can be derived and shown to lead to a stable behavior while guaranteeing the feasibility at all times. Our solution strategy is based on reformulating the original problem as the unconstrained optimization of a loss function proposed in [13] and identifying the conditions under which the loss function is convex as well as strongly convex. We show how these algorithms can be extended to the case when inequality constraints are present as well. Conditions under which stable behavior can be guaranteed are clearly delineated in all cases.

The organization of the paper is as follows. Section II outlines the problem statement and the approach we adopt to find its solution, along with a few preliminaries on convex analysis and constrained optimization. In Section III, we first propose a HT-based algorithm for equality-constrained convex optimization problems and based on that, generalize our treatment to general convex problems involving inequality constraints. We then extend our approach for convex optimization to a class of nonconvex problems in Section IV. In Section V, we summarize the main contributions of the paper and outline a few future research directions.

The first author was supported by a Siemens Fellowship. This work was supported in part by Boeing Strategic University Initiative.

A. Parashar, P. Srivastava and A. M. Annaswamy are with the Department of Mechanical Engineering, Massachusetts Institute of Technology. Email:{[anjalip](mailto:anjalip@mit.edu),[psrivast](mailto:psrivast@mit.edu),[aanna](mailto:aanna@mit.edu)}@mit.edu.

II. PROBLEM STATEMENT

We consider optimization problems of the form

$$\begin{aligned} \min f(x) \\ \text{s.t. } h(x) = 0 \\ g(x) \leq 0, \end{aligned} \quad (1)$$

where $x \in \mathbb{R}^n$ is the decision variable, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, and $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ are continuously differentiable functions. Without loss of generality, we assume that problem (1) is not overdetermined, i.e., $m \leq n$.

Our aim is to design recursive algorithms that solve (1), exhibiting accelerated convergence as proposed by Nesterov for unconstrained convex optimization, cf. [10]. Our approach is based on reformulating the original problem as the unconstrained optimization of a loss function and then using high-order tuners for accelerated convergence, cf. [11]. To satisfy constraints closely during the evolution, we employ a constraint-completion and correction procedure described below. As will become apparent, our approach will rely on ensuring that an underlying loss function is convex. We first show that this is indeed the case if the optimization problem is convex in Section III. We then generalize our treatment to a class of nonconvex optimization problems in Section IV.

Before proceeding to the technical content of the paper, we present our notational conventions and review basic concepts from convex analysis and constrained optimization below.

Notation

Let \mathbb{R} denote the set of real numbers. $\|\cdot\|$ denotes the 2-norm of a vector or matrix. For a continuously differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, ∇f denotes its gradient. $(\cdot)^T$ denotes the transpose of a vector or matrix. For vectors $x, y \in \mathbb{R}^n$, $x \geq y$ implies that the inequality holds elementwise. For a vector $x \in \mathbb{R}^n$, with $1 \leq i < j \leq n$, $x_{i:j}$ denotes the subvector with elements from the i -th entry of x to the j -th entry.

Convex analysis

Here we present the basics of convex analysis following [4], [14]. A set $\Omega \in \mathbb{R}^n$ is convex if for all $x, y \in \Omega$, $\lambda x + (1 - \lambda)y \in \Omega$ for all $\lambda \in (0, 1)$.

Definition 1. (*Convex functions*): A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex on a convex set Ω if for all $x, y \in \Omega$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y),$$

for all $\lambda \in (0, 1)$.

Definition 2. (*Strongly convex functions*): A continuously differentiable function f is μ -strongly convex on Ω if there exists a $\mu > 0$ such that for all $x, y \in \Omega$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \frac{1}{2}\mu\lambda(1 - \lambda)\|x - y\|^2,$$

for all $\lambda \in (0, 1)$.

Definition 3. (\bar{L} -smooth convex functions): A continuously differentiable function f is \bar{L} -smooth convex on Ω if it is

convex and there exists an upper bound $\bar{L} > 0$ on the Lipschitz constant of its gradient, i.e.,

$$\frac{\|\nabla f(x) - \nabla f(y)\|}{\|x - y\|} \leq \bar{L},$$

for all $x, y \in \Omega$.

A similar definition could be stated for \bar{L} -smooth strongly convex functions but we omit it here to avoid repetition.

Nonlinear optimization via constraint-completion and correction

Consider the nonlinear optimization problem (1) and define a loss function $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}$, consisting of the original objective function and soft loss terms penalizing the constraint violation as

$$\mathcal{L}(x) = f(x) + \lambda_h \|h(x)\|^2 + \lambda_g \|\text{softplus}(g(x))\|^2,$$

where $\lambda_h, \lambda_g > 0$ are design parameters. Given $y \in \mathbb{R}^p$, the function $\text{softplus} : \mathbb{R}^p \rightarrow \mathbb{R}^p$ is defined as

$$\text{softplus}(y) = \log(1 + e^y),$$

and serves as a smooth approximation to ReLU to ensure that \mathcal{L} is continuously differentiable. We employ a constraint-completion and correction approach to leverage the fact that equality constraints introduce linear dependencies in the feasible solution space, as in [13]. Building on this insight and assuming that problem (1) is not overdetermined, x is partitioned into an independent variable $\theta \in \mathbb{R}^m$ and a dependent variable $z \in \mathbb{R}^{n-m}$,

$$x = [\theta^T \quad z^T]^T.$$

We assume that h is such that given m entries of x , its remaining $(n - m)$ entries can be computed either in closed form or recursively. In other words, we assume that we have knowledge of the function $p : \mathbb{R}^m \rightarrow \mathbb{R}^{n-m}$ such that

$$h([\theta^T \quad p(\theta)^T]^T) = 0,$$

holds for all $\theta \in \mathbb{R}^m$. For all the points where $\frac{\partial h}{\partial z} \neq 0$, existence and uniqueness of p is guaranteed from the Implicit Function theorem. The reduction of variable dimension as explained above ensures that equality constraints are always satisfied.

Using the function $p(\cdot)$ defined as above, we now define a modified loss function $l : \mathbb{R}^m \rightarrow \mathbb{R}$ as

$$l(\theta) = \mathcal{L}([\theta^T \quad p(\theta)^T]^T).$$

The optimization problem in (1) is now reformulated as an unconstrained minimization problem given by

$$\min l(\theta),$$

with $\theta \in \mathbb{R}^m$ as the decision variable. Depending on the information about the mapping p , gradient of the modified loss function could be computed either explicitly or using the Implicit Function theorem as in [15].

The above completion procedure takes care of the equality constraints. However, there is no guarantee associated with the satisfaction of inequality constraints. As will be shown later, we employ a gradient-based correction procedure that employs penalty terms using the softplus function above and corrections that allow the solution to approach the feasible region along the manifold of points satisfying the equality constraints.

III. CONVEX OPTIMIZATION PROBLEMS

In this section, we consider convex optimization problems in the general form

$$\begin{aligned} \min f(x) \\ \text{s.t. } Ax = b \\ g(x) \leq 0, \end{aligned} \quad (2)$$

where $x \in \mathbb{R}^n$ is the decision variable, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ are continuously differentiable (strongly) convex functions, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. We start with problems involving just the equality constraints and then generalize our approach to problems involving both equality and inequality constraints.

A. Equality-constrained convex optimization

Equality-constrained convex optimization problems have the general structure

$$\begin{aligned} \min f(x) \\ \text{s.t. } Ax = b. \end{aligned} \quad (3)$$

With $\lambda_h > 0$, the loss function \mathcal{L} and the modified loss function l take the following forms:

$$\mathcal{L}(x) = f(x) + \lambda_h \|Ax - b\|^2, \quad (4a)$$

$$l(\theta) = f([\theta^T \ p(\theta)^T]^T). \quad (4b)$$

Here we have used the fact that $\lambda_h \|A[\theta^T \ p(\theta)^T]^T - b\|^2 = 0$. We note that p is an affine function of θ . Let

$$p(\theta) = P\theta + q, \quad (5)$$

where $P \in \mathbb{R}^{(n-m) \times m}$ and $q \in \mathbb{R}^{n-m}$. \mathcal{L} is convex by construction. We characterize the convexity properties of l in the following result.

Proposition III.1. (Convexity of the modified loss function for equality-constrained convex programs): For the equality-constrained convex optimization problem (3), assume f is a \bar{L} -smooth convex function, and let

$$\bar{M} = \sqrt{1 + \|P\|} \bar{L}. \quad (6)$$

Then l is \bar{M} -smooth convex.

Proof. Convexity of l follows in a straightforward manner from the definitions of l and p in (4b) and (5) respectively, and the convexity of f . For the \bar{M} -smoothness, using the chain rule, we have

$$\frac{dl}{d\theta} = \frac{\partial l}{\partial \theta} + \frac{\partial l}{\partial p(\theta)} \frac{\partial p(\theta)}{\partial \theta}.$$

Hence

$$\nabla l = \nabla f_{1:m} + P^T \nabla f_{m+1:n}.$$

Let us consider the gradient of l at $\theta_1, \theta_2 \in \mathbb{R}^m$ and examine the Lipschitz constant of l .

$$\begin{aligned} \frac{\|\nabla l(\theta_1) - \nabla l(\theta_2)\|}{\|\theta_1 - \theta_2\|} &\leq \frac{\|\nabla f_{1:m}(\theta_1) - \nabla f_{1:m}(\theta_2)\|}{\|\theta_1 - \theta_2\|} \\ &\quad + \frac{\|P^T(\nabla f_{m+1:n}(\theta_1) - \nabla f_{m+1:n}(\theta_2))\|}{\|\theta_1 - \theta_2\|} \end{aligned}$$

Since $\|\nabla f_{i:j}\| \leq \|\nabla f\|$ for all i, j , we have

$$\frac{\|\nabla l(\theta_1) - \nabla l(\theta_2)\|}{\|\theta_1 - \theta_2\|} \leq \frac{(1 + \|P\|)\|\nabla f(x_1) - \nabla f(x_2)\|}{\|\theta_1 - \theta_2\|} \quad (7)$$

To write the denominator of the above expression in terms of x_1 and x_2 , remember that

$$\|\theta_1 - \theta_2\|^2 + \|P\theta_1 - P\theta_2\|^2 = \|x_1 - x_2\|^2.$$

Using properties of the norm,

$$\begin{aligned} \|\theta_1 - \theta_2\|^2 + \|P\|\|\theta_1 - P\theta_2\|^2 &\geq \|x_1 - x_2\|^2 \\ \|\theta_1 - \theta_2\| &\geq \frac{\|x_1 - x_2\|}{\sqrt{1 + \|P\|}}. \end{aligned} \quad (8)$$

Using (7) and (8), we have

$$\begin{aligned} \frac{\|\nabla l(\theta_1) - \nabla l(\theta_2)\|}{\|\theta_1 - \theta_2\|} &\leq \frac{\sqrt{1 + \|P\|}\|\nabla f(x_1) - \nabla f(x_2)\|}{\|x_1 - x_2\|} \\ &\leq \sqrt{1 + \|P\|} \bar{L}, \end{aligned}$$

where the last inequality follows from the \bar{L} -smoothness property of f . \square

The next result extends Proposition III.1 to the case of strongly convex functions.

Corollary III.2. (Strong convexity of the modified loss function for equality-constrained convex programs): For the equality-constrained convex optimization problem (3), assume f is a \bar{L} -smooth and μ -strongly convex function. Then l is \bar{M} -smooth and μ -strongly convex, where \bar{M} is defined in equation (6).

Proof. Consider $\theta_1, \theta_2 \in \mathbb{R}^m$. Now consider $x_1, x_2 \in \mathbb{R}^n$ defined as $x_1 = [\theta_1^T \ p(\theta_1)^T]^T$ and $x_2 = [\theta_2^T \ p(\theta_2)^T]^T$. Then from the properties of f , it follows that

$$\begin{aligned} f(\lambda x_1 + (1 - \lambda)x_2) &\leq \lambda f(x_1) + (1 - \lambda)f(x_2) \\ &\quad - \frac{1}{2}\mu\lambda(1 - \lambda)\|x_1 - x_2\|^2, \end{aligned}$$

for all $\lambda \in (0, 1)$. Using the fact that $\|\theta_1 - \theta_2\|^2 \leq \|x_1 - x_2\|^2$, we have

$$\begin{aligned} f(\lambda x_1 + (1 - \lambda)x_2) &\leq \lambda f(x_1) + (1 - \lambda)f(x_2) \\ &\quad - \frac{1}{2}\mu\lambda(1 - \lambda)\|\theta_1 - \theta_2\|^2. \end{aligned}$$

Rest of the proof follows from the definition of l and the proof of Proposition III.1. \square

Now that we have established the convexity and smoothness properties of the modified loss function l , we leverage the properties of high-order tuners [6], [11] to propose an accelerated algorithm to solve (3). Let \mathcal{N}_k be the normalizing signal defined as

$$\mathcal{N}_k = 1 + H_k,$$

where

$$H_k = \max\{\zeta : \zeta \in \sigma(\nabla^2 l(\theta_k))\},$$

where $\sigma(\nabla^2 l(\theta_k))$ denotes the spectrum of the Hessian matrix of the loss function l evaluated at $\theta = \theta_k$. Note that it is also possible to make a more conservative selection for \mathcal{N}_k such as \bar{M} , i.e., smoothness parameter of the loss function if accurate information about $\nabla^2 l$ is not available. Next we introduce Algorithm 1 to solve problem (3).

Algorithm 1 HT Optimizer for equality-constrained convex optimization

- 1: **Initial conditions** θ_0, ν_0 , gains γ, β
 - 2: **for** $k = 0$ to N **do**
 - 3: Compute $\nabla l(\theta)$ and let $\mathcal{N}_k = 1 + H_k$
 - 4: $\nabla \bar{q}_k(\theta_k) = \frac{\nabla l(\theta_k)}{\mathcal{N}_k}$
 - 5: $\bar{\theta}_k = \theta_k - \gamma\beta \nabla \bar{q}_k(\theta_k)$
 - 6: $\theta_{k+1} \leftarrow \bar{\theta}_k - \beta(\bar{\theta}_k - \nu_k)$
 - 7: Compute $\nabla l(\theta_{k+1})$ and let
 - 8: $\nabla \bar{q}_k(\theta_{k+1}) = \frac{\nabla l(\theta_{k+1})}{\mathcal{N}_k}$
 - 9: $\nu_{k+1} \leftarrow \nu_k - \gamma \nabla \bar{q}_k(\theta_{k+1})$
 - 10: **end for**
-

The following result formally characterizes the convergence properties of Algorithm 1.

Theorem III.3. (Convergence of the HT algorithm for equality-constrained convex programs): *If the objective function f is \bar{L} -smooth convex, then with $0 < \beta < 1$ and $0 < \gamma < \frac{\beta(2-\beta)}{8+\beta}$, the sequence of iterates $\{\theta_k\}$ generated by Algorithm 1 satisfy $\lim_{k \rightarrow \infty} l(\theta_k) = l(\theta^*)$, where $l(\theta^*) = f([\theta^{*T} \ p(\theta^*)^T]^T)$ is the optimal value of (3).*

Proof. From Proposition III.1, \bar{L} -smoothness of the objective function f implies that the loss function l is \bar{M} -smooth and convex. Rest of the proof follows from [11, Theorem 2]. \square

Following Corollary III.2 and [11, Theorem 3], a similar result exists for the strongly convex case as well.

B. Convex optimization problems with equality & inequality constraints

Here we extend our approach to solve general convex optimization problems involving equality as well as the inequality constraints in the form (2). Redefine the loss function and the modified loss functions (4) by including a penalty term corresponding to the inequality constraints

violation as

$$\mathcal{L}(x) = f(x) + \lambda_h \|Ax - b\|^2 + \lambda_g \|\text{softplus}(g(x))\|^2, \quad (9a)$$

$$l(\theta) = f([\theta^T \ p(\theta)^T]^T) + \lambda_g \|\text{softplus}(g([\theta^T \ p(\theta)^T]^T))\|^2. \quad (9b)$$

Since the penalty term corresponding to the inequality constraints is convex, it follows from Proposition III.1 that the modified loss function (9b) is convex.

Our approach is based on the inequality correction procedure of [13], briefly described in Section II. The method involves first implementing the HT Algorithm 1 on the loss function (9b) ensuring that the equality constraints are met at all times. Then we apply an additional update that drives the decision variable towards the feasible region corresponding to the inequality constraints as well. Let $\alpha > 0$ be the stepsize and define $\rho : \mathbb{R}^n \rightarrow \mathbb{R}^n$ as

$$\rho \left(\begin{bmatrix} \theta \\ p(\theta) \end{bmatrix} \right) = \begin{bmatrix} \theta - \alpha \Delta \theta \\ p(\theta) - \alpha \Delta p(\theta) \end{bmatrix}, \quad (10)$$

where

$$\Delta \theta = \left(\frac{d}{d\theta} \|\text{softplus} \left(g \left(\begin{bmatrix} \theta \\ p(\theta) \end{bmatrix} \right) \right) \|^2 \right)^T.$$

Note immediately that the inequality correction step above does not affect the feasibility with respect to the equality constraints. Hence by implementing the described method, we obtain Algorithm 2 that satisfies equality constraints at each step and moves closer towards satisfying the inequality constraints with each successive iteration.

Algorithm 2 HT optimizer for equality + inequality constrained convex optimization

- 1: **Initial conditions** x_0, \bar{x}_0, ν_0 , gains α, γ, β
 - 2: **for** $k=0,1,2,\dots$ **do**
 - 3: Compute $\nabla l(\theta_k)$ and let $\mathcal{N}_k = 1 + H_k$
 - 4: $\nabla \bar{f}_k(\theta_k) = \frac{\nabla l(\theta_k)}{\mathcal{N}_k}$
 - 5: $\bar{\theta}_k = \theta_k - \gamma\beta \nabla \bar{f}_k(\theta_k)$
 - 6: $\theta_{k+1} \leftarrow \bar{\theta}_k - \beta(\bar{\theta}_k - \nu_k)$
 - 7: $\bar{x}_{k+1} = [\theta_{k+1}^T \ p(\theta_{k+1})^T]^T$
 - 8: Compute $x_{k+1} \leftarrow \rho(\bar{x}_{k+1})$
 - 9: Compute $\nabla l(\theta_{k+1})$ and let
 - 10: $\nabla \bar{f}_k(\theta_{k+1}) = \frac{\nabla l(\theta_{k+1})}{\mathcal{N}_k}$
 - 11: $\nu_{k+1} \leftarrow \nu_k - \gamma \nabla \bar{f}_k(\theta_{k+1})$
 - 12: **end for**
-

It is reasonable to expect that if the hypotheses of Theorem III.3 are satisfied and α is properly selected, Algorithm 2 solves problem (2).

IV. CONVEX OPTIMIZATION FOR A CLASS OF NONCONVEX PROBLEMS

In this section, we extend our approach of achieving accelerated convergence via high-order tuners to a class of

nonconvex optimization problems. As with the convex case, we start with problems involving just the equality constraints and then generalize our methodology to problems involving inequality constraints as well. Throughout this section, we consider only the conditions under which the loss function is convex. The arguments could be generalized easily to the strongly convex case.

A. Equality-constrained nonconvex problems

Consider the optimization problem

$$\begin{aligned} \min f(x) \\ \text{s.t. } h(x) = 0, \end{aligned} \quad (11)$$

and the associated loss function

$$\mathcal{L}(x) = f(x) + \lambda_h \|h(x)\|^2. \quad (12)$$

The following result whose proof is immediate and hence skipped, provides the conditions under which \mathcal{L} is convex.

Lemma IV.1. (Convexity of the loss function for equality-constrained nonconvex programs): *If f and h are convex, then \mathcal{L} defined in (12) is convex.*

The modified loss function in this case once again takes the form (4b), albeit the functional form of p would not be linear anymore. As such, establishing the convexity of l over the entire domain as in Proposition III.1 may not be feasible anymore. We therefore search for conditions under which the modified loss function l is convex over some subset of the domain. We summarize a set of such conditions in the following result.

Proposition IV.2. (Convexity of the modified loss function for equality-constrained nonconvex programs): *Assume that there exists a convex set $\Omega_n \in \mathbb{R}^n$ such that the functions f and h are convex on Ω_n . Let*

$$\Omega_m = \{\theta \mid \theta = x_{1:m}, x \in \Omega_n\}. \quad (13)$$

If either of the following conditions is satisfied:

- (i) $\nabla \mathcal{L}(x) \geq 0$ for all $x \in \Omega_n$, and p is convex on Ω_m ,
 - (ii) $\nabla \mathcal{L}(x) \leq 0$ for all $x \in \Omega_n$, and p is concave on Ω_m ,
- then l is convex on Ω_m .

Proof. We present here the arguments for only condition (i). The ensuing treatment easily generalizes to condition (ii). It is immediate to see that Ω_m is convex. Consider $\theta_1, \theta_2 \in \Omega_m$. Since $p(\theta)$ is convex, we have

$$p(\lambda\theta_1 + (1-\lambda)\theta_2) \leq \lambda p(\theta_1) + (1-\lambda)p(\theta_2),$$

and it follows that

$$\begin{bmatrix} \lambda\theta_1 + (1-\lambda)\theta_2 \\ p(\lambda\theta_1 + (1-\lambda)\theta_2) \end{bmatrix} \leq \lambda \begin{bmatrix} \theta_1 \\ p(\theta_1) \end{bmatrix} + (1-\lambda) \begin{bmatrix} \theta_2 \\ p(\theta_2) \end{bmatrix}.$$

Since \mathcal{L} is nondecreasing, it follows that

$$\begin{aligned} & \mathcal{L} \left(\begin{bmatrix} \lambda\theta_1 + (1-\lambda)\theta_2 \\ p(\lambda\theta_1 + (1-\lambda)\theta_2) \end{bmatrix} \right) \\ & \leq \mathcal{L} \left(\lambda \begin{bmatrix} \theta_1 \\ p(\theta_1) \end{bmatrix} + (1-\lambda) \begin{bmatrix} \theta_2 \\ p(\theta_2) \end{bmatrix} \right). \end{aligned} \quad (14)$$

Moreover, it follows from Lemma IV.1 that \mathcal{L} is convex, and we have

$$\begin{aligned} & \mathcal{L} \left(\lambda \begin{bmatrix} \theta_1 \\ p(\theta_1) \end{bmatrix} + (1-\lambda) \begin{bmatrix} \theta_2 \\ p(\theta_2) \end{bmatrix} \right) \\ & \leq \lambda \mathcal{L} \left(\begin{bmatrix} \theta_1 \\ p(\theta_1) \end{bmatrix} \right) + (1-\lambda) \mathcal{L} \left(\begin{bmatrix} \theta_2 \\ p(\theta_2) \end{bmatrix} \right). \end{aligned} \quad (15)$$

Combining the inequalities (14) and (15), we get

$$\begin{aligned} & \mathcal{L} \left(\begin{bmatrix} \lambda\theta_1 + (1-\lambda)\theta_2 \\ p(\lambda\theta_1 + (1-\lambda)\theta_2) \end{bmatrix} \right) \\ & \leq \lambda \mathcal{L} \left(\begin{bmatrix} \theta_1 \\ p(\theta_1) \end{bmatrix} \right) + (1-\lambda) \mathcal{L} \left(\begin{bmatrix} \theta_2 \\ p(\theta_2) \end{bmatrix} \right). \end{aligned}$$

And from the definition of the modified loss function, it follows that

$$l(\lambda\theta_1 + (1-\lambda)\theta_2) \leq \lambda l(\theta_1) + (1-\lambda)l(\theta_2),$$

completing the proof. \square

Now that we have established sufficient conditions for the convexity of the modified loss function, we can use high-order tuners to find an optimizer of (11). In fact, if the sequence of iterates lie within the set Ω_n , then we can use Algorithm 1, stated earlier for convex programs, to find a solution of (11). The following result formalizes this.

Theorem IV.3. (Convergence of the HT algorithm for equality-constrained nonconvex programs): *If the objective function f and the equality constraint h are convex over a set Ω_n , In addition, with $0 < \beta < 1$, $0 < \gamma < \frac{\beta(2-\beta)}{8+\beta}$, and $\theta_0 \in \Omega_m$, where Ω_m is defined in (13), if the sequence of iterates $\{\theta_k\}$ generated by Algorithm 1 satisfy $\{\theta_k\} \in \Omega_m$, then $\lim_{k \rightarrow \infty} l(\theta_k) = l(\theta^*)$, where $l(\theta^*) = f([\theta^{*T} \ p(\theta^*)^T]^T)$ is the optimal value of (11).*

Proof. Convexity of the loss function l follows from Proposition IV.2. Moreover, since $\{\theta_k\} \in \Omega_m$ for all k , there exists a constant \bar{S} such that

$$\frac{\|\nabla l(\theta_1) - \nabla l(\theta_2)\|}{\|\theta_1 - \theta_2\|} \leq \bar{S},$$

for all $\theta_1, \theta_2 \in \Omega_m$. Rest of the proof follows from [11, Theorem 2]. \square

Theorem IV.3 enables us to leverage Algorithm 1, provided that the state remains inside the set over which the modified loss function is convex. It is reasonable to argue that this is always not the case. To overcome this assumption, we use the projection operator defined as

$$\text{proj}_{\Omega_m}(\tilde{\theta}) = \text{argmin} \|\tilde{\theta} - \theta\|, \quad \forall \theta \in \Omega_m$$

to make sure that the state remains inside the set Ω_m . Algorithm 3 states this concisely.

The arguments of this section show how the proposed approach could be applied to solve nonconvex problems, where a convex objective function needs to be optimized with

Algorithm 3 HT Optimizer for equality-constrained nonconvex optimization

```

1: Initial conditions  $\theta_0, \nu_0$ , gains  $\gamma, \beta$ 
2:  $\theta_0 \leftarrow \text{proj}_{\Omega_m}(\theta_0)$ 
3: for  $k = 1$  to  $N$  do
4:   Compute  $\nabla l(\theta)$  and let  $\mathcal{N}_k = 1 + H_k$ 
5:    $\nabla \bar{q}_k(\theta_k) = \frac{\nabla l(\theta_k)}{\mathcal{N}_k}$ 
6:    $\bar{\theta}_k = \theta_k - \gamma\beta\nabla \bar{q}_k(\theta_k)$ 
7:    $\theta_{k+1} \leftarrow \text{proj}_{\Omega_m}(\bar{\theta}_k - \beta(\bar{\theta}_k - \nu_k))$ 
8:   Compute  $\nabla l(\theta_{k+1})$  and let
9:    $\nabla \bar{q}_k(\theta_{k+1}) = \frac{\nabla l(\theta_{k+1})}{\mathcal{N}_k}$ 
10:   $\nu_{k+1} \leftarrow \nu_k - \gamma\nabla \bar{q}_k(\theta_{k+1})$ 
11: end for

```

respect to nonlinear convex equality constraints. Moreover, the proposed method ensures that the state remains feasible throughout the evolution, regardless of the convexity of the original problem.

B. Nonconvex optimization problems with equality & inequality constraints

Here we employ the correction procedure described in Section II to extend our approach to solve problems of the form (1), which includes inequality constraints in addition to equality constraints. As in the convex case, let us define the function ρ as in (10) and consider the following algorithm, obtained from appending Algorithm 3 with a suitable inequality correction step.

Algorithm 4 HT optimizer for equality + inequality constrained nonconvex optimization

```

1: Initial conditions  $x_0, \bar{x}_0, \nu_0$ , gains  $\alpha, \gamma, \beta$ 
2:  $x_0 \leftarrow \text{proj}_{\Omega_n}(x_0)$ 
3: for  $k=0,1,2,\dots$  do
4:   Compute  $\nabla l(\theta_k)$  and let  $\mathcal{N}_k = 1 + H_k$ 
5:    $\nabla \bar{f}_k(\theta_k) = \frac{\nabla l(\theta_k)}{\mathcal{N}_k}$ 
6:    $\bar{\theta}_k = \theta_k - \gamma\beta\nabla \bar{f}_k(\theta_k)$ 
7:    $\theta_{k+1} \leftarrow \bar{\theta}_k - \beta(\bar{\theta}_k - \nu_k)$ 
8:    $\bar{x}_{k+1} = [\theta_{k+1}^T \ p(\theta_{k+1})^T]^T$ 
9:   Compute  $x_{k+1} \leftarrow \text{proj}_{\Omega_n}(\rho(\bar{x}_{k+1}))$ 
10:  Compute  $\nabla l(\theta_{k+1})$  and let
11:   $\nabla \bar{f}_k(\theta_{k+1}) = \frac{\nabla l(\theta_{k+1})}{\mathcal{N}_k}$ 
12:   $\nu_{k+1} \leftarrow \nu_k - \gamma\nabla \bar{q}_k(\theta_{k+1})$ 
13: end for

```

If the hypotheses of Theorem IV.3 are satisfied, then with an appropriate value of α introduced in (10), we can use Algorithm 4 to find a solution of (1), without the explicit assumption of states belonging to the set Ω_n .

V. CONCLUSIONS AND FUTURE WORK

We have presented accelerated algorithms based on high-order tuners for solving constrained convex optimization problems. Our approach is based on identifying the conditions under which the reformulated loss function is convex, guarantees that the equality constraints are satisfied at all times, and is also applicable to a class of nonconvex optimization problems. Future work will involve formally characterizing the rate of convergence and extending our approach to a broader class of nonconvex problems.

REFERENCES

- [1] K. S. Narendra and A. M. Annaswamy, *Stable Adaptive Systems*. NJ: Prentice-Hall, Inc., 1989, (out of print).
- [2] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.
- [3] E. Hazan, "Introduction to online convex optimization," *arXiv preprint arXiv:1909.05207*, 2019.
- [4] Y. Nesterov, *Lectures on convex optimization*. Springer International Publishing, 2018, vol. 137.
- [5] J. E. Gaudio, A. M. Annaswamy, M. A. Bolender, E. Lavretsky, and T. E. Gibson, "A Class of High Order Tuners for Adaptive Systems," *IEEE Control Systems Letters*, pp. 1–1, 6 2020.
- [6] J. E. Gaudio, A. M. Annaswamy, J. M. Moreu, M. A. Bolender, and T. E. Gibson, "Accelerated Learning with Robustness to Adversarial Regressors," *3rd LADC Conference*, 2021.
- [7] N. M. Boffi and J.-J. E. Slotine, "Implicit regularization and momentum algorithms in nonlinear adaptive control and prediction," *arXiv preprint arXiv:1912.13154*, 2020.
- [8] A. S. Morse, "High-order parameter tuners for the adaptive control of linear and nonlinear systems," in *Systems, models and feedback: Theory and Applications*. Springer, 1992, pp. 339–364.
- [9] S. Evesque, A. Annaswamy, S. Niculescu, and A. Dowling, "Adaptive control of a class of time-delay systems," *J. Dyn. Sys., Meas., Control*, vol. 125, no. 2, pp. 186–193, 2003.
- [10] Y. E. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.
- [11] J. M. Moreu and A. M. Annaswamy, "A stable high-order tuner for general convex functions," *IEEE Control Systems Letters*, vol. 6, pp. 566–571, 2022.
- [12] P. Srivastava and J. Cortés, "Nesterov acceleration for equality-constrained convex optimization via continuously differentiable penalty functions," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 415–420, 2021.
- [13] P. L. Donti, D. Rolnick, and J. Z. Kolter, "DC3: A learning method for optimization with hard constraints," in *International Conference on Learning Representations*, 2021.
- [14] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton University Press, 1970.
- [15] B. Amos and J. Z. Kolter, "OptNet: Differentiable optimization as a layer in neural networks," in *International Conference on Machine Learning*, 2017.