Time-Varying Optimization with Optimal Parametric Functions

Mohsen Amidzadeh, *Member, IEEE*
Department of Computer Science, Aalto University, Finland
mohsen.amidzade@aalto.fi

Abstract—In this paper, we consider a formulation of non-linear constrained optimization problems. We reformulate it as a time-varying optimization using continuous-time parametric functions and derive a dynamical system for tracking the optimal solution. We then re-parameterize the dynamical system to express it based on a linear combination of the parametric functions. Calculus of variations is applied to optimize the parametric functions, so that the optimality distance of the solution is minimized. Accordingly, an iterative dynamic algorithm, named as OP-TVO, is devised to find the solution with an efficient convergence rate. We benchmark the performance of the proposed algorithm with the prediction-correction method (PCM) from the optimality and computational complexity point-of-views. The results show that OP-TVO can compete with PCM for the optimization problem of interest, which indicates it can be a promising approach to replace PCM for some time-varying optimization problems. Furthermore, this work provides a novel paradigm for solving parametric dynamical system.

Index Terms—Time-varying optimization problem, functional optimization problem, prediction-correction method, optimality distance, dynamical system.

I. INTRODUCTION

Time-Varying Optimization (TVO) problems pertain to parametric optimization problems when the objective, constraints, or both are parameterized using continuously-varying functions. This paradigm is exploited to find the optimal trajectory of solutions in continuous-time optimization challenges. In addition, for optimization problems for which the optimal solution is known for one specific configuration, TVO can be used to extrapolate the solution to the settings of interest.

TVO is studied in the context of parametric programming [1]–[5], where the optimization problem is parametrized using continuous parameters. In [3], the authors develop prediction-correction methods to solve nonlinear constrained TVO problems. Their approach tracks a trajectory path with some convergence guarantees. A path-following procedure has been presented in [4] to trace a solution path of a parametric nonlinear problem. The authors utilize quadratic programming as a tracking procedure and derive some convergence properties for their approach. In [5], the authors leverage a path-following method to track the solution of parametric nonlinear constrained programs using a semi-smooth barrier function.

TVO can be also considered as an extension of time-invariant optimization problems [6]–[10]. In [6], a discrete-time prediction-correction approach has been proposed to minimize unconstrained time-varying functions. They analyze the asymptotic tracking error to ensure convergence of the solution. In [7], the authors present prediction-correction methods to track the optimal solution trajectory, in the primal space, with a bounded asymptotical error. Then, in [8], an interior-point method is developed for optimization problems with time-varying cost and constraint functions. The authors formulate a continuous-time dynamical system to track the optimal solution with asymptotical tracking error being bounded. In [9], prediction-correction methods have been devised, in the dual space, to track the solutions of time-varying linearly constrained problems.

When it comes to the approach for solving TVO problems, the prediction-correction schemes as promising tracking algorithms, should be addressed [6], [7], [9], [11]–[14]. It constitutes a dynamic-tracking procedure or predictor step to trace the solution trajectory over time, and a Newton-based iterative method or corrector step to adjust the error of predictions. Substantial efforts have been made to analyze the convergence of prediction-correction approaches and ensure the boundedness of tracking error [7], [8], [11].

In this paper, we study a class of nonlinear constrained optimization problems whose optimal solution is known for a setting, and the optimal solution is aimed for a target configuration. We then exploit the notion of TVO to reformulate the problem based on parametric programming. We then explore a set of parametric functions such that the optimality distance is optimized. In contrast to the works based on the prediction-correction methods, we focus on a functional optimization problem to expedite the convergence rate. In other words, this paper differs from prediction-correction approaches, in the sense that the parametric functions are designed to minimize the optimality distance of the solution instead of correcting the predictions.

The main contributions of this paper are listed as follows:

- We re-express a class of nonlinear constrained optimization problems as a time-varying optimization problem with continuous-time parametric functions. We then use

a re-parametrization trick to represent the corresponding dynamical system based on a linear combination of the parametric functions.

- We then globally optimize the parametric functions using a functional optimization problem, and devise an iterative algorithm with the optimality distance of the solutions being minimized. We call the devised algorithm Optimal Parametric Time-Varying Optimization (OP-TVO).
- We compare OP-TVO with a prediction-correction method from the literature, from the optimality and computational complexity perspectives.

Notations: In this paper, we use lower-case a for scalars, bold-face lower-case $\mathbf{a}$ for vectors and bold-face uppercase $\mathbf{A}$ for matrices. Further, $\mathbf{A}^\top$ is the transpose of $\mathbf{A}$, $\|\mathbf{a}\|$ is the euclidean norm of $\mathbf{a}$, $\nabla_{\mathbf{a}} g(\cdot)$ and $\nabla_{\mathbf{a}}^2 g(\cdot)$ are the gradient vector and Hessian matrix of multivariate function $g(\mathbf{a})$ with respect to (w.r.t.) vector $\mathbf{a}$, respectively. We show the components of a n-dimensional column vector $\boldsymbol{a}$ using the notation $\boldsymbol{a} = [a_1, \ldots, a_n]^\top$. Further, $\{a_n\}_1^N$ collects the components of vector $\boldsymbol{a}$ from $n = 1$ to $n = N$. We use $\mathbf{I}$, $\mathbf{1}$ and $\mathbf{0}$ to denote the identity matrix, all-ones and all-zeros vectors, respectively. We use $\dot{\boldsymbol{a}}(\theta)$ to represent the derivative of $\boldsymbol{a}(\theta)$ w.r.t. θ.

II. PROBLEM STATEMENT

This paper considers a class of nonlinear constrained optimization problems. The problem includes an objective function $f(\cdot): \mathbb{R}^N \to \mathbb{R}$, vector-valued constraint functions $\boldsymbol{h}_m(\cdot) = [h_{m,1}, \ldots, h_{m,N}]^\top(\cdot): \mathbb{R}^N \to \mathbb{R}^N$ and optimization variables $\boldsymbol{x}_m \in \mathbb{R}^N$ for $m \in \{1, \ldots, M\}$. The problem under study is:

$$P_0: \quad \min_{\{\boldsymbol{x}_m\}_1^M} \sum_{m=1}^M a_m f(\boldsymbol{x_m})$$
$$\text{s.t.} \quad \sum_{m=1}^M \boldsymbol{h}_m(\boldsymbol{x}_m) = \boldsymbol{u}, \quad (1)$$

where $\boldsymbol{u} \in \mathbb{R}^N$. We assume that the optimal solution of Problem P_0 is known for the given non-zero parameters $a_m = p_{0,m}$, with $m \in \{1, \ldots, M\}$. However, the optimal solution is unknown and desired for the non-zero target parameters $a_m = p_{\tau,m}$.

Note that this type of optimization problem arises in distributed optimizations or multi-agent systems where distinct agents produce agent-specific rewards and together make an overall cost function. The aim thus is obtaining an agent-specific variables $\boldsymbol{x}_m$ optimizing this overall cost function. This also may arise in constrained problems with an objective being established from different cost functions with distinct weights $\{a_m\}_1^M$.

The goal is to express P_0 based on a TVO problem with parametric functions, and devise a path-following method with convergence rate being optimized. To optimize it, we follow a functional optimization approach to design the parametric functions. As such, we parameterize $\{a_m\}_1^M$ with the parametric functions $\{b_m(\theta)\}_1^M$ and parameter $\theta \in \mathbb{R}^+ \cup \{0\}$, so that

$$\lim_{\theta \to 0} b_m(\theta) = p_{0,m}, \qquad \lim_{\theta \to \tau} b_m(\theta) = p_{\tau,m}, \quad (2)$$

for $m \in \{1, \ldots, M\}$. Note that such parametric functions as presented in (2) are not unique and we aim to find the optimal function throughout this work.

We then consider the following TVO problem

$$P_1(\theta): \qquad \boldsymbol{x}^*(\theta) = \operatorname*{argmin}_{\{\boldsymbol{x}_m(\theta)\}_1^M} \sum_{m=1}^M b_m(\theta) f\big(\boldsymbol{x_m}(\theta)\big)$$
$$\text{s.t.} \quad \sum_{m=1}^M \boldsymbol{h}_m\big(\boldsymbol{x}_m(\theta)\big) = \boldsymbol{u}, \quad (3)$$

where $\boldsymbol{x}(\theta) = [\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_M^\top]^\top(\theta)$. As declared, we assume that the solution of $P_1(\theta)$ for $\theta = 0$ is given, and the solution at target $\theta = \tau > 0$ is to be found.

A naive approach to find the solution of $P_1(\theta)$ is to use a Newton-based iterative algorithm. However, this approach suffers from low convergence rate and its solution optimality depends on a step-length parameter.

Instead, we develop a time-varying approach to exploit the information of optimal solution of $P_1(0)$. We thus formulate a continuous-time dynamical system

$$\dot{\boldsymbol{x}}_m(\theta) = \boldsymbol{\phi}_m\big(\boldsymbol{x}(\theta), \theta\big): \; \mathbb{R}^{NM} \times \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^N, \quad (4)$$

with optimal trajectory solution denoted by $\boldsymbol{x}^*(\cdot)$. We then devise an iterative approach to predict $\boldsymbol{x}^*(\cdot)$ by $\boldsymbol{x}(\cdot)$ so that the optimality distance $\|\boldsymbol{x}(\theta) - \boldsymbol{x}^*(\theta)\|$ is minimized for $\theta \to \tau$. Note that (4) shows a set of ODEs which should be solved with initial condition $\boldsymbol{x}(0)$ to give the target point $\boldsymbol{x}(\tau)$. As such, we intend to jointly solve TVO $P_1(\theta)$ and design $\{b_m(\theta)\}_1^M$ for $\theta \in [0, \tau]$ such that the optimality distance is optimized.

Let the following assumptions hold for TVO $P_1(\theta)$.

Assumption I: The objective function $f(\cdot)$ and constraints $\boldsymbol{h}_m(\cdot)$ are twice continuously differentiable with respect to (w.r.t.) $\boldsymbol{x}_m$.

Assumption II: The matrices $\{\boldsymbol{\mathcal{J}}_m\}_1^M$ are invertible for $\theta \in [0, \tau]$, where $\boldsymbol{\mathcal{J}}_m \in \mathbb{R}^{N \times N}$ is the transpose of Jacobian matrix of $\boldsymbol{h}_m(\cdot)$ w.r.t. $\boldsymbol{x}_m(\theta)$.

III. ODES ASSOCIATED WITH TVO $P_1(\theta)$

Proposition 1. *The solution of Karush–Kuhn–Tucker conditions of problem $P_1(\theta)$, for $\theta \in [0, \tau]$, can be found by the pair $(\boldsymbol{x}(\theta), \boldsymbol{\lambda})$ which follows the dynamical system (4) with:*

$$\boldsymbol{\phi}_m\big(\boldsymbol{x}(\theta), \theta\big) =$$
$$- \bigg(b_m(\theta)\nabla_m^2 f + \sum_{n=1}^N \lambda_n \nabla_m^2 h_{m,n}\bigg)^{-1} \boldsymbol{\mathcal{J}}_m \bigg(\dot{\boldsymbol{\lambda}} - \frac{\dot{b}_m(\theta)}{b_m(\theta)}\boldsymbol{\lambda}\bigg),$$
$$(5)$$

where

$$\boldsymbol{\lambda} = -b_m(\theta)\boldsymbol{\mathcal{J}}_m^{-1}\nabla_m f, \qquad \text{for} \;\; m \in \{1, \ldots, M\}, \quad (6)$$

and $\dot{\boldsymbol{\lambda}}$ is obtained using

$$\sum_{m=1}^{M} \boldsymbol{\mathcal{J}}_m^\top \dot{\boldsymbol{x}}_m(\theta) = \mathbf{0}. \tag{7}$$

Proof. Please refer to the pre-print version [15]. □

According to (5), the dynamical system $\{\boldsymbol{\phi}_m(\cdot, \cdot)\}_1^M$ has been formulated based on a nonlinear combination of two parametric functions, $b_m(\theta)$ and $\dot{b}_m(\theta)$. In the sequel, we use a decomposition trick to express the dynamical system as a linear combination of parametric functions.

A. Reparameterizing based on a Decomposition

We introduce the parametric functions

$$c_m(\theta) := \frac{\dot{b}_m(\theta)}{b_m(\theta)}, \qquad m \in \{1, \ldots, M\}, \tag{8}$$

which should satisfy $\int_0^\tau c_m(\theta)d\theta = \log\left(\frac{p_{\tau,m}}{p_{0,m}}\right) := \psi_m$ based on (2). We then have:

Theorem 1. *The dynamical system (5) can be re-parameterized based on the following linear combination of $\{c_m(\theta)\}_1^M$:*

$$\dot{\boldsymbol{x}}(\theta) = \boldsymbol{\phi}\big(\boldsymbol{x}(\theta), \theta\big) = \boldsymbol{\Gamma}(\theta)\, \boldsymbol{c}(\theta), \tag{9}$$

where $\boldsymbol{\phi}(\cdot, \cdot) = [\boldsymbol{\phi}_1^\top, \ldots, \boldsymbol{\phi}_M^\top]^\top(\cdot, \cdot)$,

$$\boldsymbol{\Gamma}(\theta) = \begin{pmatrix} \boldsymbol{\gamma}_{11} & \cdots & \boldsymbol{\gamma}_{1M} \\ \vdots & \vdots & \vdots \\ \boldsymbol{\gamma}_{N1} & \cdots & \boldsymbol{\gamma}_{NM} \end{pmatrix} \in \mathbb{R}^{NM \times M},$$

$$\boldsymbol{\gamma}_{nm} = -G_n^{-1}\left(\left(\sum_{k=1}^{M} D_k\right)^{-1} D_m - \delta_{nm}\boldsymbol{I}\right)\mathbf{1},$$

$$G_n = \operatorname{diag}(\boldsymbol{v})^{-1}\boldsymbol{\mathcal{J}}_n^{-1}\left(\nabla_n^2 f + \sum_{j=1}^{N} v_j \nabla_n^2 h_{n,j}\right),$$

$$D_m = \boldsymbol{\mathcal{J}}_m^\top G_m^{-1},$$

for $n \in \{1, \ldots, N\}$ and $m \in \{1, \ldots, M\}$ with

$$\boldsymbol{v} = -\boldsymbol{\mathcal{J}}_m^{-1} \nabla_m f. \tag{10}$$

Proof. We introduce N-by-N matrices $\{G_m\}_1^M$ and exploit the following decomposition:

$$b_m(\theta)\nabla_m^2 f + \sum_{n=1}^{N} \lambda_n \nabla_m^2 h_{m,n} = \boldsymbol{\mathcal{J}}_m \operatorname{diag}(\boldsymbol{\lambda})G_m. \tag{11}$$

Then, we get:

$$G_m = \operatorname{diag}(\boldsymbol{v})^{-1}\boldsymbol{\mathcal{J}}_m^{-1}\left(\nabla_m^2 f + \sum_{n=1}^{N} v_n \nabla_m^2 h_{m,n}\right), \tag{12}$$

for which we used $\boldsymbol{\lambda} = b_m(\theta)\boldsymbol{v}$ based on (6) and (10). Equation (12) shows that G_m is notably independent of parametric function $b_n(\theta)$. Now, by plugging (11) into (5), we obtain:

$$\dot{\boldsymbol{x}}_n(\theta) = -G_n^{-1}\operatorname{diag}(\boldsymbol{\lambda})^{-1}\big(\dot{\boldsymbol{\lambda}} - c_n(\theta)\boldsymbol{\lambda}\big),$$

and according to (7), we get:

$$\dot{\boldsymbol{\lambda}} = \boldsymbol{\lambda}\left(\sum_{k=1}^{M} D_k\right)^{-1}\sum_{m=1}^{M} D_m c_m(\theta)\,\mathbf{1},$$

which together yields:

$$\dot{\boldsymbol{x}}_n(\theta) = -G_n^{-1}\left(\left(\sum_{k=1}^{M} D_k\right)^{-1}\sum_{m=1}^{M} D_m c_m(\theta) - c_n(\theta)\boldsymbol{I}\right)\mathbf{1}.$$

Based on the definition of $\boldsymbol{\gamma}_{nm}$, it thus reads:

$$\dot{\boldsymbol{x}}_n(\theta) = \sum_{m=1}^{M} \boldsymbol{\gamma}_{nm} c_m(\theta), \quad n \in \{1, \ldots, N\},$$

which proves the statement. □

Note that the dynamical system (9) depends only on $\boldsymbol{c}(\cdot)$ and not on $\boldsymbol{b}(\cdot)$. Moreover, considering that $\boldsymbol{\Gamma}(\cdot)$ does not depend on $\boldsymbol{c}(\cdot)$, the dynamical system portrays a linear parametric expression w.r.t. $\boldsymbol{c}(\cdot)$. It enables us to find the condition in which solving (5) or equivalently (9) leads to a unique solution. Let us first make the following assumptions.

Assumption III: The matrices $\{\nabla_n^2(f + \boldsymbol{v}^\top \boldsymbol{h}_n)\}_1^N$, $\sum_{k=1}^{M} D_k$ and $\operatorname{diag}(\boldsymbol{v})$ are invertible.

Assumption IV: The derivatives of $\{\nabla_m f\}_1^M$, $\{\boldsymbol{\mathcal{J}}_m\}_1^M$ and $\{\nabla_n^2(f + \boldsymbol{v}^\top \boldsymbol{h}_n)\}_1^N$ w.r.t. $\{\boldsymbol{x}_m(\theta)\}_1^M$ are bounded.

Proposition 2. *If Assumptions I, II, III and IV hold, then the dynamical system (9) has a unique solution.*

Proof. By applying Picard–Lindelöf theorem and considering the equivalence between Lipschitz continuity and derivative boundedness, the statement follows. □

The linear form of (9) also enables us to design $\boldsymbol{c}(\cdot)$ such that the optimality distance is optimized. Having $\boldsymbol{c}(\cdot)$ being designed, we can get: $b_m(\theta) = p_{0,m} \exp\big(\int_0^\theta c_m(\xi)d\xi\big)$ based on (2) and (8).

IV. OPTIMALITY DISTANCE

Equation (9) shows a set of ODEs that is intricate to precisely solve due to highly non-linearity w.r.t. θ. However, one approach is to use the Euler method [16] to approximate $\boldsymbol{x}(\cdot)$ by $\hat{\boldsymbol{x}}(\cdot)$ based on a sequential manner:

$$\hat{\boldsymbol{x}}(\theta) = \hat{\boldsymbol{x}}(\theta - \Delta\theta) + \Delta\theta\,\dot{\hat{\boldsymbol{x}}}(\theta - \Delta\theta), \qquad \theta \in (0, \tau], \tag{13}$$

where $\Delta\theta$ is the incremental step and $\dot{\hat{\boldsymbol{x}}}(\theta) = \boldsymbol{\phi}\big(\hat{\boldsymbol{x}}(\theta), \theta\big)$. For this method, the optimality distance O_d is upper-bounded by:

$$O_d = \|\hat{\boldsymbol{x}}(\tau) - \boldsymbol{x}(\tau)\| \leq \frac{\Delta\theta^2}{2}\sum_{j=1}^{L} \|\ddot{\boldsymbol{x}}(\tau - j\Delta\theta)\| + \mathcal{O}(L\Delta\theta^3),$$

where $\tau = L\Delta\theta$. This shows that optimality distance is limited by order of $\Delta\theta^2$. However, based on (13) another upper-bound can be found as follows:

$$O_d = \left\|\int_0^\tau \big(\dot{\boldsymbol{x}}(\theta) - \hat{\boldsymbol{m}}\big)d\theta\right\| \leq \int_0^\tau \|\dot{\boldsymbol{x}}(\theta) - \hat{\boldsymbol{m}}\|d\theta, \tag{14}$$

where

$$\hat{m} = \frac{1}{L} \sum_{j=1}^{L} \hat{x}(\tau - j\Delta\theta).$$

Note that $\hat{m}$ does not depend on θ. Consequently, minimizing the upper-bound of O_d, i.e., $\int_0^\tau \|\dot{x}(\theta) - \hat{m}\| d\theta$, leads to the optimality distance being minimized.

V. OPTIMALITY DISTANCE MINIMIZATION

We consider the following functional optimization problem (FOP) to jointly design the parametric functions $c(\cdot)$ and find the optimum solution $x(\cdot)$:

$$
\begin{aligned}
J_1: \quad & \min_{x(\cdot), c(\cdot)} \int_0^\tau \|\dot{x}(\theta) - \hat{m}\|^2 d\theta \\
& \text{s.t.} \quad \dot{x}(\theta) = \Gamma(\theta)\, c(\theta), \\
& \text{s.t.} \quad \int_0^\tau c(\theta) d\theta = \psi.
\end{aligned}
\tag{15}
$$

By solving FOP J_1, we can achieve the optimal solution $x(\theta)$ which minimizes the optimality distance O_d based on (14). To solve (15), we constitute the Hamiltonian $\mathcal{H}$ as:

$$\mathcal{H} = \|\dot{x}(\theta) - \hat{m}\|^2 + w(\theta)^\top \big(\dot{x}(\theta) - \Gamma(\theta) c(\theta)\big) + \lambda^\top \Big(c(\theta) - \frac{1}{\tau}\psi\Big),$$

where $w(\theta)$ is a co-state variables and λ is a Lagrange multiplier. Using calculus of variations, the functional solution of (15) is obtained by:

$$
\begin{cases}
\nabla_{c(\theta)}\mathcal{H} = 2\Gamma(\theta)^\top \Big(\Gamma(\theta)c(\theta) - \hat{m} - \frac{1}{2}w(\theta)\Big) + \lambda = 0, \\[2mm]
\nabla_{x(\theta)}\mathcal{H} - \frac{d}{d\theta}\nabla_{\dot{x}(\theta)}\mathcal{H} \\[1mm]
\quad = w(\theta)^\top \nabla_{x(\theta)}\Gamma(\theta)\, c(\theta) + 2\ddot{x}(\theta) + \dot{w}(\theta) = 0, \\[2mm]
\dot{x}(\theta) - \Gamma(\theta)\, c(\theta) = 0, \qquad \int_0^\tau c(\theta)d\theta - \psi = 0.
\end{cases}
\tag{16}
$$

The conditions in (16) are intricate to solve. Further, an estimation of $\hat{m}$ is needed in advance. This motivates us to devise an iterative algorithm to find the solution of J_1.

In this regard, we perform an iterative algorithm as follows: In the beginning, we initialize $c(\cdot)$ such that $\int_0^\tau c(\theta)d\theta = \psi$. Then, we repeatedly follow two consecutive steps till the algorithm converges. These are the *Prediction* and *Parametral-tuning* steps. In the prediction step of this iterative approach, we sequentially solve (13) using (9) and recently updated $c(\cdot)$, in order to predict $x(\theta)$ for $\theta \in (0, \tau]$. In the parametral-tuning step, we minimize the functional objective $\int_0^\tau \|\dot{x}(\theta) - \hat{m}\|^2 d\theta$ w.r.t. $c(\cdot)$ with $\hat{m}$ being obtained based on the solution of prediction step. As declared, we execute these two steps till the convergence. We call this algorithm Optimal Parametric Time-Varying Optimization (OP-TVO).

More specifically, we consider the following FOP, in the second step of OP-TVO, to optimize the parametric functions $c(\cdot)$:

$$
\begin{aligned}
J_2: \quad & \min_{c(\cdot)} \int_0^\tau \big\|\Gamma(\theta)\, c(\theta) - \hat{m}\big\|^2 d\theta + \mu \int_0^\tau c(\theta)^\top c(\theta) d\theta \\
& \text{s.t.} \quad \int_0^\tau c(\theta) d\theta = \psi,
\end{aligned}
\tag{17}
$$

Algorithm 1: Optimal Parametric Time-Varying Optimization (OP-TVO)

Input: Optimal solution $x(0)$.
Outputs: Optimal solution $x(\tau)$ and parametric functions $c(\cdot)$.
Initialization:
    Initialize $c(\cdot)$ so that $\int_0^\tau c(\theta)d\theta = \psi$.
    Set flag = 1 and iter = 0.
while flag **do**
    iter = iter+1.

    Prediction step:
    Predict $x(\theta)$ for $\theta \in (0, \tau]$ using (13) and (9) and based on updated $c(\cdot)$.

    Parametral-tuning step:
    Update $c(\cdot)$ using (18) and based on predicted $x(\theta)$ with $\theta \in (0, \tau]$.

    if *Convergence* **then**
    | flag = 0.
    end
end

---

where the term $\int_0^\tau c(\theta)^\top c(\theta) d\theta$ is additionally added to regulate the smoothness of $c(\theta)$ w.r.t. θ, and $0 < \mu \ll 1$ is the regulation coefficient.

Proposition 3. *The globally optimal solution of J_2 is obtained by:*

$$c(\theta) = \Pi^{-1}(\theta) \big(\Gamma(\theta)^\top \hat{m} - \lambda\big),
\tag{18}$$

where

$$\Pi(\theta) = \Gamma(\theta)^\top \Gamma(\theta) + \mu I,$$

$$\lambda = \left(\int_0^\tau \Pi^{-1}(\theta)d\theta\right)^{-1} \left(\int_0^\tau \Pi^{-1}(\theta)\Gamma(\theta)^\top d\theta\, \hat{m} - \psi\right).$$

Proof. Considering that the dynamical system (9) has been expressed based on a linear combination of parametric functions $c(\cdot)$, J_2 is a convex FOP. This implies that the globally optimal solution of J_2 can be found by applying the Euler-Lagrange equation on J_2 [17]. We thus get:

$$
\begin{cases}
\Gamma(\theta)^\top \big(\Gamma(\theta)\, c(\theta) - \hat{m}\big) + \mu c(\theta) + \lambda &= 0, \\
\int_0^\tau c(\theta)d\theta - \psi &= 0 .
\end{cases}
\tag{19}
$$

By solving (19), the statement follows. $\qquad\square$

Algorithm 1 shows the pseudo-code of OP-TVO. For each iteration (iter), the prediction and parametral-tuning steps are performed to jointly predict the optimal solution $x^*(\tau)$ and design the parametric functions $c(\cdot)$. We also need a metric for the convergence to stop the algorithm. For this, we track the value of $\hat{O}_d := \sum_{j=1}^{L} \left\|\phi\Big(\hat{x}(\tau - j\Delta\theta), \tau - j\Delta\theta\Big) - \hat{m}\right\|^2$ as an estimation of the optimality distance. We thus consider

that the algorithm has converged if the value of $\hat{O}_d$ lies below a threshold O_{th}.

VI. NUMERICAL RESULTS AND DISCUSSION

To evaluate the devised algorithm OP-TVO, we compare it with a Prediction-Correction Method (PCM) [6], as well as with a *Benchmark* solution being obtained using an extremely small incremental step $\Delta\theta = 10^{-6}$, disregarding its computational complexity. Note that, we consider this *Benchmark* as the optimal solution, by which we can compute the optimality distance O_d. These algorithms have been implemented using Matlab R2022a on a 8×1.70 GHz Intel Core i5-10310U Processor, equipped with 16 GB of memory and 12 Mbytes of data cache.

We then consider the constrained optimization problem E_1 [18], [19], and change the constraints to add non-linearity to the problem.

$$E_1: \quad \min_{\{\boldsymbol{x}_m\}_1^M} \sum_{m=1}^{M} a_m \, \text{erfc}\left(\gamma_0 \frac{x_{m,1}}{\sqrt{2^{\frac{0.1}{x_{m,2}}} - 1}} \right)$$

$$\text{s.t.} \quad \sum_{m=1}^{M} \log(1 + x_{m,1}) = L,$$

$$\text{s.t.} \quad \sum_{m=1}^{M} x_{m,2}^2 = 1,$$

where the optimization variables are $\boldsymbol{x}_m = [x_{m,1}, x_{m,2}]^\top$ for $m \in \{1, \ldots, M\}$, $M = 100$, $a_m = m^{-\tau} / \sum_{m=1}^{M} m^{-\tau}$, $\tau = 3$ and $\gamma_0 = 40$. For E_1, it can be verified that the optimal solution when $a_m = \frac{1}{M}$ is obtained as:

$$x_{m,1} = \exp(L/M - 1), \quad x_{m,2} = \sqrt{1/M}, \quad m \in \{1, \ldots, M\}.$$

Therefore, we constitute a TVO problem exactly as E_1 but with parametric functions $\{b_m(\theta)\}_1^M$ replacing $\{a_m\}_1^M$ such that:

$$b_m(0) = \frac{1}{M}, \quad b_m(\tau) = a_m, \quad m \in \{1, \ldots, M\}.$$

We further use the re-parametrizing functions $\boldsymbol{c}(\cdot)$ with $c_m(\theta) := \frac{\dot{b}_m(\theta)}{b_m(\theta)}$. We then apply Algorithm 1 with hyper-parameters $\mu = 10^{-7}$, $O_{\text{th}} = 10^{-5}$ and $\Delta\theta = 10^{-2}$ to find the optimal solution.

Figures 1 and 2 illustrate the trajectory solutions $\{\boldsymbol{x}_m(\theta)\}_1^M$, being obtained by Algorithm 1, as a function of θ for different iterations. For the first iteration, the trajectory solutions portray a highly nonlinear behavior. However, as the number of iteration increases, this non-linearity reduces. When the algorithm converges (3rd iteration), the linear curvature of the trajectory solutions indicates that a precise solution with minimal optimality distance has been obtained.

We also apply a PCM on Problem E_1 to obtain the solution. To have a fair comparison, we adjust the hyper-parameter $\Delta\theta$ for PCM such that the corresponding optimal
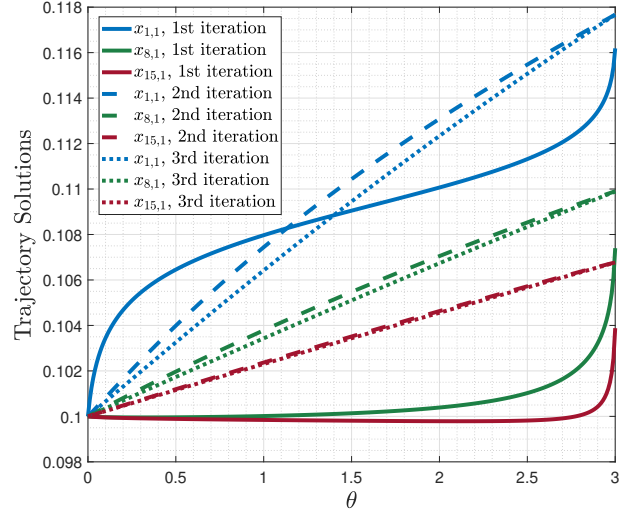

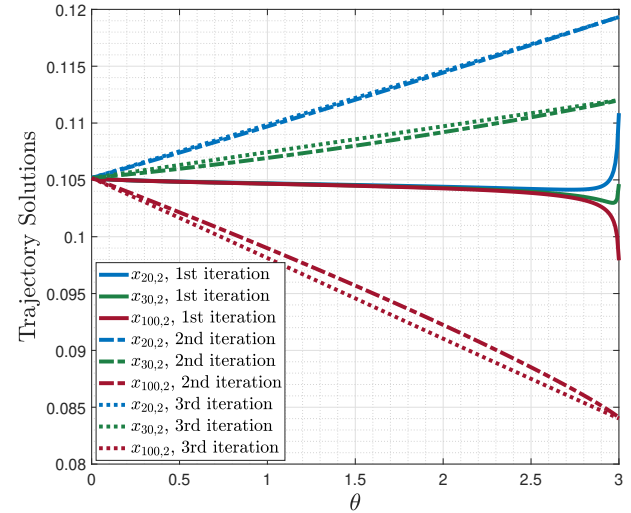Fig. 1: Trajectory Solutions for $x_{m,1}$ with $m \in \{1, 8, 15\}$.

Fig. 2: Trajectory Solutions for $x_{m,2}$ with $m \in \{20, 30, 100\}$.

value is almost equal to that of Algorithm 1. As such, we need to decrease it by 1×10^{-4}.

Table I compares the performance results of PCM, Benchmark and OP-TVO. The second column shows the optimal values that these approaches achieve, The third column states the extent to which these methods violate the constraints, represented as the summation of constraint violations, the fourth column indicates the elapsed time, in seconds, and the fifth and sixth columns show the optimality distance O_d and its estimation $\hat{O}_d$, respectively.

Although PCM achieves a slightly lower objective function value, it is important to highlight that Benchmark notably excels in terms of constraint satisfaction compared to PCM. As such, the Benchmark solution is regarded as the reference approach. According to the values of $\hat{O}_d$, OP-TVO converges after three iterations. It can be seen that OP-TVO with iter $= 3$ outperforms PCM from the computational

TABLE I: Performance Result of OP-TVO and PCM.

Approach	Optimal Value	Constraint Violations	Elapsed Time [s]	O_d	$\hat{O}_d$
`Benchmark`	5.6089×10^{-7}	1.312×10^{-6}	2066	0.0	N/A
PCM	5.6086×10^{-7}	1.739×10^{-4}	271	0.00135	N/A
OP-TVO, iter=1	3.2443×10^{-5}	0.419	3	0.8291	N/A
iter=2	5.6088×10^{-7}	5.129×10^{-5}	52	5.132×10^{-4}	0.228
iter=3	5.6088×10^{-7}	5.051×10^{-5}	105	5.130×10^{-4}	2.68×10^{-6}
iter=4	5.6088×10^{-7}	5.050×10^{-5}	160	5.130×10^{-4}	2.14×10^{-6}

complexity perspective as it converges within 105 seconds while PCM converges after 271 seconds. Furthermore, the optimality distance O_d of OP-TVO with $\text{iter} = 3$ is superior to the optimality distance of PCM. Not to mention that OP-TVO better satisfies the constraints than PCM. These imply that OP-TVO gives a more accurate solution compared to PCM, and depict that OP-TVO can find the optimal solution with lower computational complexity and better performance precision.

VII. CONCLUSION

In this paper, we reformulated a class of nonlinear constrained optimization problems based on a time-varying optimization with some parametric functions. We then leveraged a re-parametrization trick to find a dynamical system being linearly expressed in terms of the parametric functions. To minimize the optimality distance of the solution, being traced by a dynamical system, we utilized a functional minimization problem. As such, an iterative algorithm, called OP-TVO, was devised to find the solution trajectory with optimal optimality distance. Based on the results, the proposed algorithm outperforms the Prediction-Correction Method (PCM) from the optimality distance and convergence rate perspectives. The results show that OP-TVO can be considered as a promising approach to replace PCM for distributed time-varying optimization problems. Optimization problems with time-varying constraints can be considered as a future work.

VIII. ACKNOWLEDGEMENT

REFERENCES

[1] J. Guddat, F. G. Vazquez, and H. T. Jongen, *Parametric Optimization: Singularities, Pathfollowing and Jumps.* Chichester, U.K.: Wiley, 1990.

[2] Q. T. Dinh, C. Savorgnan, and M. Diehl, "Adjoint-based predictor-corrector sequential convex programming for parametric nonlinear optimization," *SIAM J. Optim.*, vol. 22, no. 4, p. 1258–1284, 2012.

[3] V. Kungurtsev and M. Diehl, "Sequential quadratic programming methods for parametric nonlinear optimization," *Computational Optimization and Applications,*, vol. 59, no. 3, pp. 475–509, Feb. 2014.

[4] V. Kungurtsev and J. Jäschke, "A predictor-corrector path-following algorithm for dual-degenerate parametric optimization problems," *SIAM J. Optim.*, vol. 27, no. 1, p. 538–564, 2017.

[5] D. Liao-McPherson, M. M. Nicotra, and I. V. Kolmanovsky, "A semismooth predictor corrector method for real-time constrained parametric optimization with applications in model predictive control," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 3600–3607.

[6] S. Andrea, M. Aryan, K. Alec, L. Geert, and A. Ribeiro, "A class of prediction-correction methods for time-varying convex optimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4576–4591, 2016.

[7] A. Simonetto and E. Dall'Anese, "Prediction-correction algorithms for time-varying constrained optimization," *IEEE Transactions on Signal Processing*, vol. 65, no. 20, pp. 5481–5494, 2017.

[8] M. Fazlyab, S. Paternain, V. M. Preciado, and A. Ribeiro, "Prediction-correction interior-point method for time-varying convex optimization," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 1973–1986, 2018.

[9] A. Simonetto, "Dual prediction–correction methods for linearly constrained time-varying convex programs," *IEEE Transactions on Automatic Control*, vol. 64, no. 8, pp. 3355–3361, 2019.

[10] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, "Time-varying convex optimization: Time-structured algorithms and applications," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2032–2048, 2020.

[11] D. Guo, X. Lin, Z. Su, S. Sun, and Z. Huang, "Design and analysis of two discrete-time ZD algorithms for time-varying nonlinear minimization," *Numer. Algorithms*, vol. 77, no. 1, p. 23–36, 2018.

[12] S. Paternain, M. Morari, and A. Ribeiro, "Real-time model predictive control based on prediction-correction algorithms," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 5285–5291.

[13] N. Bastianello, A. Simonetto, and R. Carli, "Prediction-correction splittings for nonsmooth time-varying optimization," in *2019 18th European Control Conference (ECC)*, 2019, pp. 1963–1968.

[14] A. Bernstein, E. Dall'Anese, and A. Simonetto, "Online primal-dual methods with measurement feedback for time-varying convex optimization," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 1978–1991, 2019.

[15] M. Amidzadeh, "Time-varying optimization with optimal parametric function," preprint ArXiv 2210.00931, 2023.

[16] A. Iserles, *A First Course in the Numerical Analysis of Differential Equations*, 2nd ed., ser. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2008.

[17] A. Seierstad and K. Sydsaeter, "Sufficient conditions in optimal control theory," *International Economic Review*, vol. 18, no. 2, pp. 367–391, 1977.

[18] M. Amidzade, H. Al-Tous, O. Tirkkonen, and G. Caire, "Cellular network caching based on multipoint multicast transmissions," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[19] ——, "Cellular traffic offloading with optimized compound single-point unicast and cache-based multipoint multicast," in *IEEE Wireless Commun. and Net. Conf. (WCNC)*, 2022, pp. 2268–2273.