2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Iteratively Preconditioned Gradient-Descent Approach for Moving Horizon Estimation Problems

Tianchen Liu[1], Kushal Chakrabarti[2], Nikhil Chopra[1]

Abstract— **Moving horizon estimation (MHE) is a state estimation method that has been extensively studied. The state estimates for the MHE problem are obtained by solving an approximation nonlinear optimization problem. This optimization process is known to be computationally challenging. This paper explores the idea of iteratively preconditioned gradient-descent (IPG) to solve the MHE issue to outperform the current solution methods in light of this limitation. To our knowledge, the preconditioning technique is employed for the first time in this research to speed up the critical MHE optimization stage and lower the computing cost. For a class of MHE problems, the proposed iterative approach's convergence guarantee is shown. Sufficient conditions for the MHE problem to be convex are also derived. Finally, the proposed method is implemented on a unicycle localization example. The simulation results demonstrate that the proposed approach can improve accuracy with reduced computational costs.**

I. Introduction

Moving horizon estimation (MHE) is an optimization-based technique for state estimation problems. It formulates and solves an optimization problem at each sampling instant to obtain the best state estimate. While using complete prior information for the estimation should generate better estimates, the computational cost can also quickly become intractable. MHE handles this challenge by utilizing a finite number of past measurements and control inputs and discarding the previous information to maintain a feasible computational cost. Compared with other estimators, like extended Kalman filter (EKF) [1], MHE performs well for the constrained state estimation problem when the arrival cost is accurately approximated, which contains information on the discarded data. A general introduction and some applications of MHE can be found in [2]. Due to its performance and efficiency, MHE has become a widely used approach for state estimation in many applications [3], [4], [5], [6]. Stability analysis has also been investigated for specific scenarios (e.g., [7], [8]).

The performance of MHE critically relies on the algorithm used to solve the underlying optimization problem [9]. Various strategies have been developed to reduce computational complexity while maintaining accuracy. In [10], Nesterov's fast gradient method expedites the optimization step but is limited to linear systems. In [11], nonlinear system equations are approximated by Carleman linearization expressions to reduce the computational cost for gradient and Hessian. In [12], three approaches based on the gradient, conjugate gradient, and Newton's method have been proposed to reduce the computational effort and demonstrated to be more effective than the Kalman filter by simulation. In this work, we introduce a preconditioning matrix [13] which is updated iteratively. Specifically, we transform the distributed iteratively preconditioned gradient-descent (IPG) approach in [14] to its centralized counterpart and employ it for the nonlinear state estimation in the MHE framework. Our approach can be deployed to solve MHE for a general nonlinear system without linear approximations. Compared with [12], where a finite number of iterations is used to reduce the computational cost, a complete optimization is fulfilled until the state estimates converge at each time-instant in our approach.

The convergence proof of the proposed algorithm and sufficient conditions for the MHE problem to be convex are presented. This is the first attempt to accelerate MHE optimization via a preconditioning technique and demonstrate a convexity analysis for MHE problems. The proposed approach is implemented on a numerical example for estimating the locations of a mobile robot. The results demonstrate that the MHE approach achieves better performance than EKF, invariant EKF (InEKF) [15], and a recently developed IPG observer [16]. Compared with the default solver in Matlab, the proposed approach can obtain the same results with a reduced computational cost. The main contributions of this paper are summarized as follows,

- To accelerate the optimization step, an algorithm using an iterative preconditioning technique [14] is developed to solve MHE problems in Section III-A. The convergence proof of the proposed algorithm is presented for convex MHE problems in Section III-B.
- Section III-C derives sufficient conditions for the convexity of MHE problems to guarantee convergence.
- Section IV validates the proposed approach to a mobile robot localization problem.

[1] Department of Mechanical Engineering, University of Maryland, College Park, MD 20742, USA. {tianchen, nchopra}@umd.edu

[2] Division of Data & Decision Sciences, Tata Consultancy Services Research, Mumbai, 400607 India. chakrabarti.k@tcs.com

*This work is supported by USDA NIFA Sustainable Agricultural Systems (SAS) Program (Award Number: 20206801231805)

II. PROBLEM DESCRIPTION

For $i = 0, \dots, T-1$, we consider the system

$$x_{i+1} = f(x_i, u_i) + w_i,\ y_i = h(x_i) + v_i, \tag{1}$$

where $x_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$, and $y_i \in \mathbb{R}^p$ denote the states, the inputs, and the observations at i^{th} sampling instant, respectively, and T is the total number of sampling steps. The process disturbance set $\mathbb{W} \subseteq \mathbb{R}^n$ and measurement noise set $\mathbb{V} \subseteq \mathbb{R}^p$ are assumed to be compact with $0 \in \mathbb{W}$ and $0 \in \mathbb{V}$ [17]. Hence, $w_i \in \mathbb{W}$ and $v_i \in \mathbb{V}$ are bounded process disturbance and measurement noise. The system drift function $f : (\mathbb{R}^n, \mathbb{R}^m) \to \mathbb{R}^n$ and the measurement function $h : \mathbb{R}^n \to \mathbb{R}^p$ are assumed to be known. The formulation indicates that an input control and a (partial) measurement occur at each sampling step.

For each time instant $t = N, \dots, T$, the MHE problem can be formulated as [18],

$$\min_{x_{\{t-N:t\}}} \Phi_{t-N} := \sum_{i=t-N}^{t-1} \left(w_i^T Q^{-1} w_i + v_i^T R^{-1} v_i \right) + \Gamma(x_{t-N}),$$
$$\text{s.t. } w_i = x_{i+1} - f(x_i, u_i),\ v_i = y_i - h(x_i). \tag{2}$$

The optimization variables are x_{t-N} to x_t, given the past N known control inputs and observations. $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{p\times p}$ are diagonal positive definite weighting matrices for the disturbances. The arrival cost $\Gamma(x_{t-N})$ summarizes the discarded past information. We employ an EKF-based approximation of the arrival cost [17],

$$\Gamma(x_{t-N}) = (x_{t-N} - \hat{x}_{t-N})^T \Pi_{(t-N)}^{-1} (x_{t-N} - \hat{x}_{t-N}) + \Phi^*_{t-N},$$

where $\hat{x}_{t-N}$ and Φ^*_{t-N} are the estimates of x_{t-N} and the optimal objective function value obtained at the previous time instant. $\Pi_{(t-N)} \in \mathbb{R}^{n\times n}$ is a positive definite weighting matrix, updated for the next time instant via the following matrix Riccati equation [17],

$$S_2 = J_f \Pi_{(t-N)} J_h^T (J_h \Pi_{(t-N)} J_h^T + R)^{-1} J_h \Pi_{(t-N)} J_f^T,$$
$$\Pi_{(t-N+1)} = J_f \Pi_{(t-N)} J_f^T - S_2 + Q, \tag{3}$$

where $J_f \in \mathbb{R}^{n\times n}$ and $J_h \in \mathbb{R}^{p\times n}$ are the Jacobian of f and h with respect to states x evaluated using x_{t-N}.

To present our results, we require a few more notations. For each $t \geq N$, $Y^{(t)} \in \mathbb{R}^{Np}$ and $U^{(t)} \in \mathbb{R}^{Nm}$ denote the concatenated column vectors of the past N consecutive measurements and control inputs before t^{th} time instant, respectively, i.e., $Y^{(t)} = \left[y_{t-N}^T, \dots, y_{t-1}^T\right]^T$ and $U^{(t)} = \left[u_{t-N}^T, \dots, u_{t-1}^T\right]^T$. We let $\|\cdot\|$, $\lambda_{max}[\cdot]$, and $\lambda_{min}[\cdot]$ denote the induced 2-norm, the largest, and smallest eigenvalue of a matrix.

Since the input and output data are known in MHE problems, constraints may not be necessary for state estimation. Hence, we focus on unconstrained MHE problems. Nevertheless, the algorithm will be extended in our future work to solve constrained MHE problems.

III. PROPOSED APPROACH

This section details the proposed IPG approach to solve the MHE problem, henceforth referred to as MHE-IPG. The convergence proof of MHE-IPG and a convexity analysis for MHE problems are presented.

A. MHE-IPG Approach

The critical contribution of the proposed approach lies in utilizing a preconditioning technique to accelerate the optimization step in solving MHE problems. For time instant $t = N, \dots, T$, the MHE-IPG steps are as follows.

Step 1. For t^{th} time instant, define the optimization variable vector $\xi^{(t)} \in \mathbb{R}^{(N+1)n}$, which is the concatenating column vector of variables $x_{t-N}, \dots, x_t$, i.e.,

$$\xi^{(t)} = [x_{t-N}^T, \dots, x_t^T]^T. \tag{4}$$

Then, the MHE problem in (2) is equivalent to

$$\min_{\xi^{(t)}} F(\xi^{(t)}, U^{(t)}, Y^{(t)}) = \sum_{i=t-N}^{t-1} (x_{i+1} - f(x_i, u_i))^T Q^{-1}$$
$$(x_{i+1} - f(x_i, u_i)) + \sum_{i=t-N}^{t-1} (y_i - h(x_i))^T R^{-1} (y_i - h(x_i))$$
$$+ (x_{t-N} - \hat{x}_{t-N})^T \Pi_{(t-N)}^{-1} (x_{t-N} - \hat{x}_{t-N}). \tag{5}$$

If $t = N$, the initial state estimate $\hat{x}_0$ and a positive definite $\Pi_{(0)}$ are chosen. Otherwise $\hat{x}_{t-N}$ and $\Pi_{(t-N)}$ are obtained from estimates at the previous instant.

Step 2. This step solves the optimization problem (5) using the idea of IPG. At each iteration $k = 0, 1, \dots$, an estimate $\xi_k^{(t)}$ and a preconditioner matrix $K \in \mathbb{R}^{(N+1)n\times(N+1)n}$ are maintained. Before the iterations start, we select the positive scalar constants β, δ, and initialize $\xi_0^{(t)}, K_0$. At iteration k, the estimate and preconditioner are updated via the following equations,

$$\xi_{k+1}^{(t)} = \xi_k^{(t)} - \delta K_k g(\xi_k^{(t)}), \tag{6}$$
$$K_{k+1} = K_k - \alpha_k [(H(\xi_k^{(t)}) + \beta I) K_k - I], \tag{7}$$

until $\|\xi_{k+1}^{(t)} - \xi_k^{(t)}\| < \epsilon$, where ϵ is a small positive tolerance value. $g(\xi_k^{(t)})$ and $H(\xi_k^{(t)})$ denote the gradient and Hessian of F with respect to ξ, evaluated at $\xi = \xi_k^{(t)}$. I is the identity matrix with the same dimension as $H(\xi_k^{(t)})$. α_k is selected following condition presented later in (11). Let $\hat{\xi}^{(t)} = \xi_{k+1}^{(t)}$ and go to Step 3.

Step 3. Decompose $\hat{\xi}^{(t)}$ to the state estimates for the past N time instants as $\hat{\xi}^{(t)} = [\hat{x}_{t-N}^T, \dots, \hat{x}_t^T]^T$. Record $\hat{x}_{t-N}^T, \dots, \hat{x}_t^T$ to the MHE results of (5) and go to the next time instant to estimate $\xi^{(t+1)}$. Utilizing $\hat{\xi}^{(t)}$, $\Pi_{(t-N+1)}$ is updated via (3). We use a 'warm-start' strategy to form $\xi_0^{(t+1)}$ for next time instant:

$$\xi_0^{(t+1)} = [\hat{x}_{t-N+1}^T, \dots, \hat{x}_t^T, f(\hat{x}_t^T)]^T. \tag{8}$$

Go to Step 1 for solving (5) at $t+1$ using $\xi_0^{(t+1)}$, $\hat{x}_{t-N+1}$ and $\Pi_{(t-N+1)}$. Repeat Steps 1-3 until $t = T$. ■

For an analog of Newton's method, $\delta = 1$. Hence, we can select δ in the range $\delta \in (0.1, 1.9)$ in Step 2. β can be selected as $\beta \in (0, 1)$. The convergence of the MHE-IPG algorithm is guaranteed as long as (10), presented later in Theorem 1, holds. In practice, a 'warm-start' strategy is used to initialize ξ_0 at every time step, and a guess of $(H(\xi) + \beta I)^{-1}$ can be used to initialize K_0.

B. Convergence Analysis of MHE-IPG

We make the following assumptions to present our convergence results of the proposed approach.

Assumption 1. The system equations f and h are assumed to satisfy certain conditions such that $F(\xi)$ is convex and twice continuously differentiable, with the minimum solution(s) of (5) exist and denoted as $\xi^{(t)*} \in \Xi^{(t)*}$. For brevity, we will denote $\xi^{(t)*}$ as ξ^*.

Assumption 2. The Hessian of $F(\xi)$, denoted by $H(\xi)$, is assumed to be Lipschitz continuous with respect to the 2-norm with Lipschitz constant γ, i.e., $\|H(\xi_1) - H(\xi_2)\| \leq \gamma\|\xi_1 - \xi_2\|$, $\forall \xi_1, \xi_2 \in \mathbb{R}^{(N+1)n}$. We further assume that $\|H(\xi^*)\|$ is upper bounded as $\|H(\xi^*)\| \leq q$ for some $q \in (0, \infty)$ and $H(\xi^*)$ is non-singular at any minimum point $\xi^* \in \Xi^*$.

Assumption 3. The gradient of $F(\xi)$, denoted by $g(\xi)$, is assumed to be l-Lipschitz continuous, i.e., $\|g(\xi_1) - g(\xi_2)\| \leq l\|\xi_1 - \xi_2\|$, $\forall \xi_1, \xi_2 \in \mathbb{R}^{(N+1)n}$.

For each time instant t, we introduce the following notation. We define the 'optimal' preconditioner matrix $K^* = (H(\xi^*) + \beta I)^{-1}$. It can be concluded that K^* is well-defined with $\beta > 0$ and Assumption 1. We denote $\eta = \|K^*\| = \|(H(\xi^*) + \beta I)^{-1}\| = \frac{1}{\lambda_{min}[H(\xi^*)]+\beta}$.

For each iteration k, we define $\tilde{K}_k = K_k - K^*$, the coefficient for convergence of K_k as $\rho_k = \|I - \alpha_k(H(\xi_k) + \beta I)\|$, and the estimation error $z_k = \xi_k - \xi^*$. Let $\rho = \sup \rho_k$. If $\beta > 0$ and $0 < \alpha_k < \frac{1}{\lambda_{max}[H(\xi_k)]+\beta}$, then $\rho_k \in [0, 1), \forall k \geq 0$ (see [14], Lemma 1).

The following lemma is essential for the convergence of our proposed method.

Lemma 1. [14] For each time instant $t \geq N$, consider the IPG update (6)-(7) with parameters $\beta, \delta > 0$, $\alpha_k \in (0, \frac{1}{\lambda_{max}[H(\xi_k)]+\beta})$. Then, under Assumptions 1-3,

$$\begin{aligned}\|\tilde{K}_{k+1}\| \leq &\rho^{k+1}\|\tilde{K}_0\| + \gamma\eta(\alpha_k\|z_k\| \\ &+ \rho\alpha_{k-1}\|z_{k-1}\| + \cdots + \rho^k\alpha_0\|z_0\|). \quad (9)\end{aligned}$$

The detailed proof can be found in [14, Appendix A.3].

Next, we present the convergence result of the proposed approach for solving (5) for any $t \geq N$. Superscript (t) is dropped for brevity.

Theorem 1. Suppose that Assumptions 1-3 holds. For each time instant $t \geq N$, consider the IPG update (6)-(7) with parameters $\beta > 0$, $\delta > 0$, and $\alpha_k \in (0, \frac{1}{\lambda_{max}[H(\xi_k)]+\beta})$. Let the initial estimate ξ_0 and preconditioner matrix K_0 be selected to satisfy

$$\frac{\delta\eta\gamma}{2}\|\xi_0 - \xi^*\| + \eta\beta + \eta q|1-\delta| + \delta l\|K_0 - K^*\| \leq \frac{1}{2\mu}, \quad (10)$$

where $\mu \in (1, \frac{1}{\rho})$ and $\eta = \|K^*\|$. If

$$\alpha_k < \min\{\frac{1}{\lambda_{max}[H(\xi_k)] + \beta}, \frac{\mu^k(1-\mu\rho)}{2l(1-(\mu\rho)^{k+1})}\}, \quad (11)$$

then for $k \geq 0$, $\|\xi_{k+1} - \xi^*\| < \frac{1}{\mu}\|\xi_k - \xi^*\|$.

Proof: The proof is deferred to Appendix A. ■

Theorem 1 implies that the estimates of the IPG approach locally converge to a solution of (5) with a linear convergence rate of at least $\frac{1}{\mu}$. Provided that the conditions in Theorem 1 hold, convergence of (6)-(7) to a minimum point $\xi^{(t)*} \in \Xi^{(t)*}$ of (5) is guaranteed at each $t \geq N$. In the absence of noise, it is assumed that certain conditions exist such that, for each $t \geq N$, the MHE problem (5) has a unique solution $\Xi^{(t)*}$ (see [12], [19], etc.). By definition of $\xi^{(t)}$ in (4) and $\xi^{(t)*}$, under such assumptions, Theorem 1 guarantees convergence of $\hat{x}_t$ to the true state x_t, for $t \geq N$. In the future, we will investigate the stability of MHE-IPG subject to noise.

C. MHE Convexity Analysis

Assumption 1 requires the converted function $F(\cdot)$ in (5) to be convex and twice continuously differentiable. Hence, we present sufficient conditions on the system dynamics $f(\cdot)$ and observation function $h(\cdot)$ given known U and Y such that Assumption 1 holds.

Given $\xi^{(t)}$ defined in (4) and the MHE formulation (5), the Hessian $H(\xi^{(t)}) \in \mathbb{R}^{(N+1)n\times(N+1)n}$ with respect to $\xi^{(t)}$ is a tridiagonal block matrix (see Appendix B). We define the matrix $\overline{\overline{H}}_i \in \mathbb{R}^{2n\times 2n}$ as

$$\overline{\overline{H}}_i = \begin{bmatrix} A_{11} & -J_f^T|_{x_{t-N+i}}Q^{-1} \\ -Q^{-1}J_f|_{x_{t-N+i}} & Q^{-1} \end{bmatrix}, \quad (12)$$

$$\begin{aligned}A_{11} = &\Pi^{-1} + \tilde{J}_{f(t-N)} + V_{f(t-N)}^T\tilde{Q}\tilde{H}_{f(t-N)} \\ &+ \tilde{J}_{h(t-N)} + V_{h(t-N)}^T\tilde{R}\tilde{H}_{h(t-N)}, \quad (i=0) \\ A_{11} = &\tilde{J}_{f(t-N+i)} + V_{f(t-N+i)}^T\tilde{Q}\tilde{H}_{f(t-N+i)} + \tilde{J}_{h(t-N+i)} \\ &+ V_{h(t-N+i)}^T\tilde{R}\tilde{H}_{h(t-N+i)}, \quad (i = 1, \ldots, N-1)\end{aligned}$$

where $(\cdot)|_{x_i}$ means the expressions are evaluated at time instant i. The matrices are calculated as follows,

$$\begin{aligned}&\tilde{J}_{f(i)} = J_f^TQ^{-1}J_f|_{x_i}, \quad \tilde{J}_{h(i)} = J_h^TR^{-1}J_h|_{x_i}, \\ &V_{f(i)} = I_n \otimes (f(x_i, u_i) - x_{i+1}), \quad \tilde{Q} = I_n \otimes Q^{-1}, \\ &V_{h(i)} = I_n \otimes (h(x_i) - y_i), \quad \tilde{R} = I_n \otimes R^{-1},\end{aligned}$$

$$\tilde{H}_{f(i)} = \begin{bmatrix} \mathcal{H}_{f(1,1,1)} & \cdots & \mathcal{H}_{f(1,1,n)} \\ \vdots & & \vdots \\ \mathcal{H}_{f(n,1,1)} & \cdots & \mathcal{H}_{f(n,1,n)} \\ \vdots & & \vdots \\ \mathcal{H}_{f(1,n,1)} & \cdots & \mathcal{H}_{f(1,n,n)} \\ \vdots & & \vdots \\ \mathcal{H}_{f(n,n,1)} & \cdots & \mathcal{H}_{f(n,n,n)} \end{bmatrix}_{|x_i}, \quad \tilde{H}_{h(i)} = \begin{bmatrix} \mathcal{H}_{h(1,1,1)} & \cdots & \mathcal{H}_{h(1,1,n)} \\ \vdots & & \vdots \\ \mathcal{H}_{h(p,1,1)} & \cdots & \mathcal{H}_{h(p,1,n)} \\ \vdots & & \vdots \\ \mathcal{H}_{h(1,n,1)} & \cdots & \mathcal{H}_{h(1,n,n)} \\ \vdots & & \vdots \\ \mathcal{H}_{h(p,n,1)} & \cdots & \mathcal{H}_{h(p,n,n)} \end{bmatrix}_{|x_i},$$

where $\otimes$ denotes the Kronecker product, $\mathcal{H}_f \in \mathbb{R}^{n\times n\times n}$ and $\mathcal{H}_h \in \mathbb{R}^{p\times n\times n}$ are two 3-dimensional tensors concatenating Hessians of f and h with respect to x: $\mathcal{H}_{f(i,j,k)} = \frac{\partial^2 f_i}{\partial x_j \partial x_k}$, $\mathcal{H}_{h(i,j,k)} = \frac{\partial^2 h_i}{\partial x_j \partial x_k}$.

Theorem 2. Consider the system dynamics function f and the observation function h in (1). If $\overline{\overline{H}}_i$, as defined in Eq. (12), is positive semi-definite for all $i = 0, \ldots, N-1$, then the MHE problem (5) for time instant t is convex.

Proof: The proof is deferred to Appendix B. ∎

Note that the arrival cost affects the convexity property via Π^{-1} in A_{11} of (12). In the EKF-based update, Π is recursively obtained from (3), which can influence the positive semi-definiteness of A_{11}. An alternative is to use a constant positive definite weighting Π. However, a slightly worse estimation accuracy is noticed in the example when using a constant Π.

IV. Experiments

In this section, we evaluate the proposed MHE-IPG approach to the localization problem of a mobile robot. The computations are performed in MATLAB 2022a on a Windows laptop with i7-9750H CPU. We use the first-order Euler discretization to convert the continuous-time unicycle kinematics into a discrete-time:

$$x_{i+1} = \begin{pmatrix} x_{i+1,1} \\ x_{i+1,2} \\ x_{i+1,3} \end{pmatrix} = \begin{pmatrix} x_{i,1} + dt \cdot u_{i,1} \cos(x_{i,3}) + \varepsilon_{i,x_1} \\ x_{i,2} + dt \cdot u_{i,1} \sin(x_{i,3}) + \varepsilon_{i,x_2} \\ x_{i,3} + dt \cdot u_{i,2} + \varepsilon_{i,x_3} \end{pmatrix}.$$

$x_{i,1}, x_{i,2}, x_{i,3}$ are the position and heading direction in the world frame coordinates. Control inputs are $u_i = [u_{i,1}, u_{i,2}]^T$, where $u_{i,1}$ is the forward speed and $u_{i,2}$ is the angular velocity. $\varepsilon_{i,x}$ is the process disturbance vector. The observations are the direct measurements of the position of the robot (e.g., global positioning system (GPS) measurements) with additive noises:

$$y_i = \begin{pmatrix} y_{i,1} \\ y_{i,2} \end{pmatrix} = \begin{pmatrix} h_1(x_i, u_i) + \varepsilon_{i,y_1} \\ h_2(x_i, u_i) + \varepsilon_{i,y_2} \end{pmatrix} = \begin{pmatrix} x_{i,1} + \varepsilon_{i,y_1} \\ x_{i,2} + \varepsilon_{i,y_2} \end{pmatrix},$$

where $\varepsilon_{i,y}$ is the measurements noise vector.

The initial states are $x_0 = [0, 0, 0]^T$. The sampling time is $dt = 0.2$, and the total number of sampling instants is $T = 200$. The control inputs are $u_i = [3, i/200]^T$ for $i = 0, ..., T-1$. The process noises $\varepsilon_{i,x_1}, \varepsilon_{i,x_2}, \varepsilon_{i,x_3} \sim N(0, 0.1)$, and the measurement noises $\varepsilon_{i,y_1}, \varepsilon_{i,y_2} \sim N(0, 0.4)$ are bounded with a maximal magnitude of 1.5. Given these parameters, it can be verified that Theorem 2 is valid for this problem.

Different nonlinear estimators have been tested for this localization problem, including EKF, invariant EKF (InEKF) [15], IPG observer [16] and the MHE approach. EKF is a widely used technique for nonlinear state estimation but may suffer from divergence. InEKF avoids the divergence issue by mapping the states to matrix Lie groups, where the converted problem is solved. The IPG observer was recently developed in [16] that uses the same iteratively preconditioning technique but in the manner of a Newton-type nonlinear observer. For the MHE approach, we use two optimization algorithms: i) BFGS method by default Matlab 'fminunc' solver ('MHE-default'), and ii) the proposed 'MHE-IPG'.

TABLE I

Error Comparison for Different Estimators

Method	Window Size (N)	Mean $\bar{e}$ (m)	Variance $\bar{e}$
Observations	-	0.4966	0.0669
EKF	1	1.0626	0.6241
InEKF	1	0.1993	0.0150
IPG Observer	5	0.2809	0.0230
MHE-default	5	0.1943	0.0099
MHE-IPG	5	0.1943	0.0099
IPG Observer	10	0.2362	0.0178
MHE-default	10	0.1935	0.0104
MHE-IPG	10	0.1935	0.0104
IPG Observer	15	0.2462	0.0193
MHE-default	15	0.1867	0.0097
MHE-IPG	15	0.1867	0.0097
IPG Observer	20	0.4116	0.0308
MHE-default	20	0.1851	0.00956
MHE-IPG	20	0.1851	0.00956

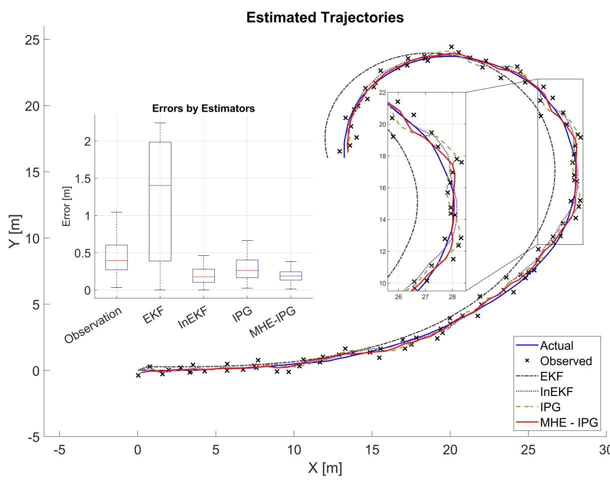

Fig. 1. Estimated Trajectories and Box Plots for Errors in One Run, with $N = 5$.

To evaluate the performance, we use the root mean square error (RMSE) over M simulation runs $\bar{e} = \frac{1}{M}\sum_{m=1}^{M}\left(\sum_{t=0}^{T}\|e_t^{(l)}\|^2\right)^{\frac{1}{2}}$, where $e_t^{(m)}$ is the estimation error of the m^{th} simulation run. $M = 30$ runs are simulated with randomly generated noises. In the results, 'Observations/Observed' refers to the metrics from raw measurement data. Table I and Fig. 1 show that all other estimators, except EKF, can obtain a lower mean and variance of RMSE than the raw measurement values. Among them, MHE-default and MHE-IPG outperform InEKF and IPG observer. As a Newton-type observer, the IPG observer tends to be more influenced by the noisy observations and thus has a slightly worse accuracy. Finally, the average error of MHE results reduces as the window size N increases, which is caused by more information being used for the estimation step.

The preconditioning technique's main benefit is accelerating the optimization step. The parameters of MHE-IPG are $\beta = 0.5$ and $\delta = 1.6$, and the same stopping

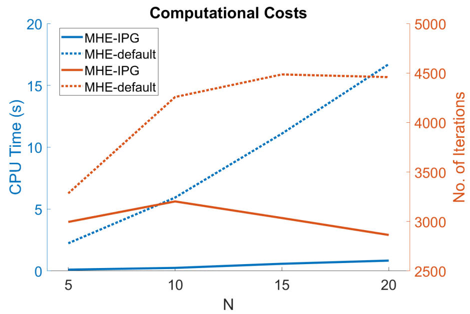

Fig. 2. Computational Cost Comparison.

criteria $\epsilon = 10^{-6}$ is used for MHE-IPG and the default 'fminunc' solver. From Fig. 2, MHE-IPG can converge in fewer iterations and run faster using the same window size N. In this example, functions for gradients and Hessians are obtained offline before the iterative estimation process to improve the computation time further. As mentioned above, a larger window size can lead to better estimation results. Hence, our proposed approach can be computationally less expensive to achieve the same level of accuracy or better with the same level of computational cost.

V. Conclusion

We proposed a new iterative approach for solving MHE problems, which utilizes a preconditioning technique to accelerate the optimization step. The convergence of the approach is rigorously analyzed, and sufficient conditions for the convexity of MHE problems are derived. Such conditions guarantee that the proposed MHE-IPG algorithm can obtain a converged estimation. Finally, the simulated unicycle localization example highlights that MHE-IPG outperforms prominent state estimators regarding accuracy. In addition, it can reduce computational costs than the default Matlab solver.

Limitations of our current work will be addressed in the future. An iterative preconditioned algorithm that can handle constrained MHE problems will be developed. In addition, we plan to employ the proposed algorithm to solve MHE problems for more complex and stiff systems and compare the results with other optimization algorithms (e.g., conjugate gradient (CG) method [20], generalized minimal residual (GMRES) method [21], etc.).

References

[1] A. Gelb, Ed., *Applied optimal estimation*. MIT press, 1974.
[2] A. Alessandri and G. Battistelli, "Moving horizon estimation: Open problems, theoretical progress, and new application perspectives," pp. 703–705, 2020.
[3] C. V. Rao, J. B. Rawlings, and J. H. Lee, "Constrained linear state estimation—a moving horizon approach," *Automatica*, vol. 37, no. 10, pp. 1619–1628, 2001.
[4] A. Alessandri, M. Baglietto, and G. Battistelli, "Receding-horizon estimation for discrete-time linear systems," *IEEE Transactions on Automatic Control*, vol. 48, no. 3, pp. 473–478, 2003.
[5] M. Osman, M. W. Mehrez, M. A. Daoud, A. Hussein, S. Jeon, and W. Melek, "A generic multi-sensor fusion scheme for localization of autonomous platforms using moving horizon estimation," *Transactions of the Institute of Measurement and Control*, vol. 43, no. 15, pp. 3413–3427, 2021.
[6] J. D. Schiller and M. A. Muller, "Suboptimal nonlinear moving horizon estimation," *IEEE Transactions on Automatic Control*, 2022.
[7] J. Liu, "Moving horizon state estimation for nonlinear systems with bounded uncertainties," *Chemical Engineering Science*, vol. 93, pp. 376–386, 2013.
[8] M. A. Müller, "Nonlinear moving horizon estimation in the presence of bounded disturbances," *Automatica*, vol. 79, pp. 306–314, 2017.
[9] L. Zou, Z. Wang, J. Hu, and Q.-L. Han, "Moving horizon estimation meets multi-sensor information fusion: Development, opportunities and challenges," *Information Fusion*, vol. 60, pp. 1–10, 2020.
[10] B. Morabito, M. Kögel, E. Bullinger, G. Pannocchia, and R. Findeisen, "Simple and efficient moving horizon estimation based on the fast gradient method," *IFAC-PapersOnLine*, vol. 48, no. 23, pp. 428–433, 2015.
[11] N. Hashemian and A. Armaou, "Fast moving horizon estimation of nonlinear processes via carleman linearization," in *2015 American Control Conference (ACC)*. IEEE, 2015, pp. 3379–3385.
[12] A. Alessandri and M. Gaggero, "Fast moving horizon state estimation for discrete-time systems using single and multi iteration descent methods," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4499–4511, 2017.
[13] C. T. Kelley, *Iterative methods for optimization*. SIAM, 1999.
[14] K. Chakrabarti, N. Gupta, and N. Chopra, "On accelerating distributed convex optimizations," *arXiv:2108.08670*, 2021.
[15] A. Barrau and S. Bonnabel, "The invariant extended kalman filter as a stable observer," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1797–1812, 2016.
[16] K. Chakrabarti and N. Chopra, "IPG observer: A newton-type observer robust to measurement noise," in *2023 American Control Conference (ACC)*. IEEE, 2023, pp. 3069–3074.
[17] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE transactions on automatic control*, vol. 48, no. 2, pp. 246–258, 2003.
[18] J. B. Rawlings and B. R. Bakshi, "Particle filtering and moving horizon estimation," *Computers & chemical engineering*, vol. 30, no. 10-12, pp. 1529–1541, 2006.
[19] A. Wynn, M. Vukov, and M. Diehl, "Convergence guarantees for moving horizon estimation based on the real-time iteration scheme," *IEEE Transactions on Automatic Control*, vol. 59, no. 8, pp. 2215–2221, 2014.
[20] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 1999.
[21] Y. Saad and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM Journal on scientific and statistical computing*, vol. 7, no. 3, pp. 856–869, 1986.

Appendix

A. Proof of Proposition 1

This proof mostly follows the proof of Theorem 1 in [14] without the assumption of $\delta = 1$. First, if we define the estimation error for the k-th iteration as $z_k = \xi_k - \xi^*$, then by (6), $z_{k+1} = \xi_{k+1} - \xi^* = \xi_k - \delta K_k g(\xi_k) - \xi^* = z_k - \delta K_k g(\xi_k)$. Given $K_k = \tilde{K}_k + K^*$, and $g(\xi^*) = 0$ by first order necessary optimality condition,

$$
\begin{aligned}
z_{k+1} &= z_k - \delta K^* g(\xi_k) - \delta \tilde{K}_k g(\xi_k) \\
&= -\delta K^*\big(g(\xi_k) - g(\xi^*) - \frac{1}{\delta}(H(\xi^*) + \beta I)z_k\big) - \delta \tilde{K}_k g(\xi_k) \\
&= -\delta K^*\Big(g(\xi_k) - g(\xi^*) - H(\xi^*)z_k\Big) + \beta K^* z_k \\
&\quad + (1-\delta)K^* H(\xi^*)z_k - \delta \tilde{K}_k g(\xi_k). \qquad (13)
\end{aligned}
$$

Nest, we find an upper bound on $\|z_{k+1}\|$. For the first term in (13), using the fundamental theorem of calculus [13], we have

$$
\begin{aligned}
&g(\xi_k) - g(\xi^*) - H(\xi^*)z_k \\
&= \int_0^1 H(s\xi_k + (1-s)\xi^*)ds(\xi_k - \xi^*) - H(\xi^*)z_k \\
&= \Big(\int_0^1 [H(s\xi_k + (1-s)\xi^*) - H(\xi^*)]ds\Big)z_k. \qquad (14)
\end{aligned}
$$

Under Assumption 2, $\|[H(s\xi_k+(1-s)\xi^*)-H(\xi^*)]\| \leq \gamma\|(s\xi_k+(1-s)\xi^*)-\xi^*)]\| = \gamma\|s(\xi_k-\xi^*)\| = \gamma s\|z_k\|$. By the definition of induced norm, (14) implies $\|g(\xi_k) - g(\xi^*) - H(\xi^*)z_k\| \leq \|z_k\|\Big(\int_0^1 \gamma s\|z_k\|ds\Big) = \frac{\gamma}{2}\|z_k\|^2$. For the last term in Eq. (13), by Assumption 3 we have $\|\delta\tilde{K}_k g(\xi_k)\| \leq \delta\|\tilde{K}_k\|\|g(\xi_k) - g(\xi^*)\| \leq \delta l\|\tilde{K}_k\|\|\xi_k - \xi^*\| = \delta l\|\tilde{K}_k\|\|z_k\|$. So, with $\eta = \|K^*\|$ and $\|H(\xi^*)\| \leq q$, (13) becomes $\|z_{k+1}\| \leq \frac{\delta\eta\gamma}{2}\|z_k\|^2 + \eta\beta\|z_k\| + \eta q|1-\delta|\|z_k\| + \delta l\|\tilde{K}_k\|\|z_k\|$. Upon substituting $\|\tilde{K}_k\|$ above from (9) in Lemma 1,

$$
\begin{aligned}
\|z_{k+1}\| &\leq \frac{\delta\eta\gamma}{2}\|z_k\|^2 + \eta q|1-\delta|\|z_k\| + \eta\beta\|z_k\| + \delta l\Big(\rho^k\|\tilde{K}_0\| \\
&\quad + \gamma\eta\alpha(\|z_{k-1}\| + \cdots + \rho^{k-1}\|z_0\|)\Big)\|z_k\|. \qquad (15)
\end{aligned}
$$

Finally, we would prove that $\|z_{k+1}\| < \frac{1}{\mu}\|z_k\|$ and $\|z_k\| < \frac{1}{\mu\delta\eta\gamma}$ are true for all k using the principle of induction. For $k = 0$, $\|z_1\| \leq \|z_0\|(\frac{\delta\eta\gamma}{2}\|z_0\| + \eta\beta + \eta q|1-\delta| + \delta l\|\tilde{K}_0\|)$. Hence, if the condition in Eq. (10) is satisfied, then $\|z_1\| \leq \frac{1}{2\mu}\|z_0\| < \frac{1}{\mu}\|z_0\|$. Also, Eq. (10) implies that $\|z_0\| < \frac{1}{\mu\delta\eta\gamma}$. Therefore, the claims are true for the first iteration.

Next, we suppose that the claims are true for the iteration 1 to iteration k. Then, $\|z_{k+1}\| < \frac{1}{\mu}\|z_k\| < \cdots < \frac{1}{\mu^{k+1}}\|z_0\| < \frac{1}{\mu^{k+1}}\frac{1}{\mu\delta\eta\gamma}$. Since $\mu > 1$, the above implies $\|z_{k+1}\| < \frac{1}{\mu\delta\eta\gamma}$. In addition, $\|z_k\| + \rho\|z_{k-1}\| + \cdots + \rho^k\|z_0\| < \|z_0\|\big(\frac{1}{\mu^k} + \frac{\rho}{\mu^{k-1}} + \cdots + \rho^k\big) = \|z_0\|\frac{1-(\mu\rho)^{k+1}}{\mu^k(1-\mu\rho)}$. For the iteration $k+1$, in order to show that $\|z_{k+2}\| < \frac{1}{\mu}\|z_{k+1}\|$, from above we have

$$
\begin{aligned}
\|z_{k+2}\| &\leq \frac{\delta\eta\gamma}{2}\|z_{k+1}\|^2 + \eta q|1-\delta|\|z_{k+1}\| + \eta\beta\|z_{k+1}\| \\
&\quad + \delta l\Big(\rho^{k+1}\|\tilde{K}_0\| + \gamma\eta\alpha(\|z_k\| + \cdots + \rho^k\|z_0\|)\Big)\|z_{k+1}\| \\
&\leq \|z_{k+1}\|\Bigg(\frac{\delta\eta\gamma}{2}\|z_{k+1}\| + \eta q|1-\delta| + \eta\beta + \delta l\rho^{k+1}\|\tilde{K}_0\| \\
&\quad + \delta l\gamma\eta\alpha\|z_0\|\frac{1-(\mu\rho)^{k+1}}{\mu^k(1-\mu\rho)}\Bigg). \qquad (16)
\end{aligned}
$$

If α is selected as $\alpha < \frac{\mu^k(1-\mu\rho)}{2l(1-(\mu\rho)^{k+1})}$, then $\delta l\gamma\eta\alpha\|z_0\|\frac{1-(\mu\rho)^{k+1}}{\mu^k(1-\mu\rho)} < \frac{\delta\eta\gamma}{2}\|z_0\|$. Since $\rho < 1$, $\delta l\rho^{k+1}\|\tilde{K}_0\| < \delta l\|\tilde{K}_0\|$. Then,

$$
\begin{aligned}
&\delta l\gamma\eta\alpha\|z_0\|\frac{1-(\mu\rho)^{k+1}}{\mu^k(1-\mu\rho)} + \eta\beta + \eta q|1-\delta| + \delta l\rho^{k+1}\|\tilde{K}_0\| \\
&< \frac{\delta\gamma\eta}{2}\|z_0\| + \eta\beta + \eta q|1-\delta| + \delta l\|\tilde{K}_0\| \leq \frac{1}{2\mu}. \qquad (17)
\end{aligned}
$$

Since $\|z_{k+1}\| < \frac{1}{\mu\delta\eta\gamma}$, upon substituting from above in (16), $\|z_{k+2}\| < \|z_{k+1}\|(\frac{1}{2\mu} + \frac{1}{2\mu}) = \frac{1}{\mu}\|z_{k+1}\|$. Hence, by the principle of induction, we have proved that $\|z_{k+1}\| < \frac{1}{\mu}\|z_k\|$ is true for all k. As $\mu > 1$, the sequence $\{\|z_k\| = \|\xi_k - \xi^*\|, \forall k\}$ is convergent.

B. Proof of Proposition 2

$H(\xi^{(t)})$ can be expressed as the sum of N matrices, as illustrated in the following form,

$$
\begin{aligned}
H(\xi^{(t)}) &= \begin{bmatrix} * & * & & & \\ * & * & * & & \\ & * & * & * & \\ & & \ddots & \ddots & \ddots \\ & & & * & * & * \\ & & & & * & * \end{bmatrix} = \begin{bmatrix} * & * & & \\ * & * & & \\ & & & \\ & & & \end{bmatrix} \\
&\quad + \begin{bmatrix} & & & \\ & * & * & \\ & * & * & \\ & & & \end{bmatrix} + \cdots + \begin{bmatrix} & & & \\ & & & \\ & & * & * \\ & & * & * \end{bmatrix} \\
&=: \overline{H}_0 + \overline{H}_1 + \cdots + \overline{H}_{N-1}. \qquad (18)
\end{aligned}
$$

The non-zero part of $\overline{H}_i$ is $\overline{\overline{H}}_i$ as defined in Eq. (12). Moreover, we use the notation $\hat{H}_{(i,j)} \in \mathbb{R}^{n\times n}$ to represent the sub-blocks of $\overline{\overline{H}}_i$,

$$
\overline{\overline{H}}_i = \begin{bmatrix} \hat{H}_{(i,i)} & \hat{H}_{(i,i+1)} \\ \hat{H}_{(i+1,i)} & \hat{H}_{(i+1,i+1)} \end{bmatrix}, \quad (i = 0, \ldots, N-1).
$$

Via tedious calculation, we can obtain that

$$
\begin{aligned}
\hat{H}_{(0,0)} &= 2\Pi^{-1} + 2\tilde{J}_{f(t-N)} + 2V_{f(t-N)}^T\tilde{Q}\tilde{H}_{f(t-N)} \\
&\quad + 2\tilde{J}_{h(t-N)} + 2V_{h(t-N)}^T\tilde{Q}\tilde{H}_{h(t-N)}, \\
\hat{H}_{(i,i)} &= 2Q^{-1} + 2\tilde{J}_{f(t-N+i)} + 2V_{f(t-N+i)}^T\tilde{Q}\tilde{H}_{f(t-N+i)} \\
&\quad + 2\tilde{J}_{h(t-N+i)} + 2V_{h(t-N+i)}^T\tilde{Q}\tilde{H}_{h(t-N+i)}, \\
&\quad (i = 1, \ldots, N-1), \quad \hat{H}_{(N,N)} = 2Q^{-1}, \\
\hat{H}_{(i,i-1)} &= -2Q^{-1}J_f|_{x_{t-N+i-1}}, \quad (i = 1, \ldots, N), \\
\hat{H}_{(i,i+1)} &= -2J_f^T|_{x_{t-N+i}}Q^{-1}, \quad (i = 0, \ldots, N-1).
\end{aligned}
$$

with the notations expressed in Section III-C. So, if $\overline{\overline{H}}_i$ is positive semi-definite, then $\overline{H}_i$ is positive semi-definite for all $i = 0, \ldots, N-1$. $H(\xi^{(t)})$ is the sum of N positive semi-definite matrices, which is also positive semi-definite. It concludes that $F(\xi^{(t)}, U^{(t)}, Y^{(t)})$ at time instant t is convex.