2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Recursive Least Squares with Log-Determinant Divergence Regularisation for Online Inertia Identification

Namhoon Cho, Taeyoon Lee, and Hyo-Sang Shin

Abstract— **This study presents a recursive algorithm for solving the regularised least squares problem for online identification of rigid body dynamic model parameters with emphasis on the physical consistency of estimated inertial parameters. One of the geometric approaches is to use a regulariser that represents how close the pseudo-inertia matrix is to a given reference on the feasible manifold in the regression problem. The proposed extension enables memory-efficient online learning in addition to the benefits of geometry-aware convex regularisation using the log-determinant divergence of the pseudo-inertia matrix. Also, the recursive version endows the estimator with the capability to deal with time-variation of parameters by introducing an optional forgetting mechanism. The characteristics of the recursive regularised least squares algorithm is demonstrated using the MIT Cheetah 3 leg swinging experiment dataset and compared to the existing batch optimisation method.**

I. Introduction

Inertia identification is a fundamental yet challenging problem in autonomy for multi-body physical systems. Examples of such systems include spacecraft with reaction wheels and appendages [1]–[5], conventional aircraft [6], unmanned aerial vehicles carrying slung-load payloads [7], humanoid and legged robots [8]–[12], and manipulators [13]–[15]. A reliable knowledge of the inertial parameters consisting of mass, centre of mass, and rotational inertia matrix is essential for precise control and efficient planning.

Inertial parameter identification is not as straightforward as it seems even though the inertial parameters appear linearly in the expression for joint torques. Technical challenges arise from both intrinsic and extrinsic sides of the problem. In the aspect of mathematical nature, physical consistency of the estimated parameters always poses a constraint that inertial parameters of a rigid body lie on certain Riemannian manifold rather than the Euclidean space. The ordinary batch or recursive least squares (RLS) with ℓ_2 regularisation often provides unrealistic results such as negative mass and negative semidefinite inertia matrix as it does not respect the non-Euclidean geometry of parameter space [16]. In the aspect of information quality, the sensor measurements are often sparse and noisy, and the operational limits imposed on the joint space usually lead to insufficient excitation in the regressor signal which renders the problem ill-posed.

Various algorithms have been developed in the recent decade to overcome the shortcomings of standard linear regression methods in inertial parameter estimation. Only a few of the algorithms are aware of the physical consistency requirement [17]. The methods can be broadly classified into two different approaches depending on how the physical feasibility of inertial parameters is enforced.

One approach is to formulate the sign definiteness requirements into linear matrix inequality (LMI) constraints and apply semidefinite programming (SDP) [1], [9], [13], [15], [18]. The LMI constraint in [13] describes positive definiteness of inertia matrix and strict positivity of mass for each body with positive definiteness of an augmented 6×6 matrix written in terms of the link inertial parameters. It was noticed in [8] that the full physical consistency requires the principal inertia moments, i.e., eigenvalues of inertia matrix, to satisfy a set of triangle inequalities in addition to the positive definiteness of inertia matrix. The approach in [1] represents the strengthened condition with a LMI constraint. The triangle inequality property is only sufficient but not necessary for the positive definiteness of inertia matrix alone, hence full physical consistency requires the solution to be confined to the corresponding subset of positive definite cone. Independent developments in [9], [15] led to the same LMI constraint given as positive definiteness of the so-called 4×4 pseudo-inertia matrix which compactly represents all physical consistency conditions.

The formulation using classical least squares criterion for fitting, standard Euclidean distance for regularisation, and LMI constraints for physical consistency turned out to have fundamental limitations. The estimates obtained as the solution to such constrained optimisation problem are often only marginally feasible. As a result, the estimated inertia distribution can be unrealistically sharp along certain directions [10], [11]. This is mainly because the optimal solution lies on the active LMI constraint boundaries when the unconstrained least squares minimiser is not physically consistent.

To address these issues, another approach focuses on developing methods based on the natural distance functions associated with the non-Euclidean geometry of the inertial parameter space [10], [11], [14], [19], [20]. The coordinate invariance of a geometric distance metric enables physically meaningful quantification of the variations in inertial parameters, which in turn manifests as better generalisability in practice. The robust offline identification method developed in [10] is to solve a nonlinear optimisation problem formulated by using the natural Riemannian metric for the manifold of physically consistent inertial parameters. Minimisation of a coordinate-invariant error criterion based on the natural Riemannian metric is theoretically appealing, however, the algebraic nonlinearity of the geodesic

Namhoon Cho and Hyo-Sang Shin are with the Centre for Autonomous and Cyber-Physical Systems, School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, United Kingdom. e-mail: {n.cho, h.shin}@cranfield.ac.uk

Taeyoon Lee is with NAVER LABS, Seongnam, Gyeonggi-do, 13561, South Korea. e-mail: ty-lee@naverlabs.com

Hyo-Sang Shin is with Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology, Daejeon, 34051, South Korea. e-mail: hyosangshin@kaist.ac.kr

979-8-3503-8457-4/24/$31.00 ©2024 IEEE

distance metric proposed in [10] lead to many difficulties in computation and theoretical convergence analysis. The Bregman divergence associated with the log-determinant of the 4×4 symmetric positive definite pseudo-inertia matrix is suggested as a surrogate function preserving coordinate invariance [11], [14]. This convex divergence metric approximates the affine-invariant Riemannian metric up to second-order. Using the Bregman divergence as a Lyapunov function yields an adaptation law that resembles natural gradient flow for more performant direct adaptive control [14]. Using the same function as a regulariser in the batch least squares framework enables efficient and robust offline identification by taking advantages of convex optimisation [11]. In [20], a singularity-free and smooth parameterisation is proposed based on the log-Cholesky decomposition of the pseudo-inertia matrix. Introducing the log-Cholesky parameterisation into the objective function of [11] leads to the unconstrained optimisation problem which is non-convex but well-posed in the sense that all local minima are global minimisers.

While the advantage of geometric regularisation in enhancing generalisation performance is clear, it has not been fully realised in an online estimation setting. The previous studies on geometric methods enabled robust identification under limited degree of observability, however, they were limited to batch optimisation. The natural adaptation law proposed in [14] performs continuous-time online update of inertial parameters on the associated manifold. However, it is a stable-tracking-oriented method for direct adaptive control which does not pursue identification but only adjusts the parameters on a need-to-know basis. Meanwhile, online inertia identification has been studied in aerospace domain, but without reference to the geometric approach. The physical consistency requirement was either incorporated with the LMI approach [1] or ignored in problem formulation [2]–[6].

This study aims to develop a recursive algorithm for online identification of inertial parameters considering physical consistency by taking advantages of geometric regularisation. Some form of online calibration or update is necessary for the estimated parameters to represent actual dynamics in the presence of uncertainties or sudden changes. The online identification capability is particularly desired in systems where the inertia of each component can change over time, e.g., spacecraft consuming fuel. In such circumstances, recursive estimation can be preferred over batch estimation as long as the theoretical well-posedness and the robust generalisability of a geometric approach can be retained.

With this background, this study takes the approach to extend the geometric method developed in [11] for batch convex optimisation to the streamed data setting. The purpose of this choice is twofold. The first is to retain the practical benefits of using the log-determinant divergence of pseudo-inertia matrix as an appropriate convex regulariser. The second is to take the own advantages of online identification including not only the memory efficiency but also the capability to estimate time-varying parameters with an adequate forgetting mechanism. The contributing points are summarised as follows:

- Development of a recursive update method for online identification of physically consistent inertial parameters based on the analytic expressions for gradient and Hessian of the log-determinant divergence
- Introduction of the optional directional forgetting mechanism for alertness to time-variation of parameters
- Verification against batch convex optimisation result using MIT Cheetah 3 fixed-base identification experiment dataset

The rest of the paper is organised as follows: Section II provides a brief overview of the geometry of inertial parameters and the log-determinant divergence. Section III presents an update method for RLS with log-determinant divergence regularisation which is essentially a Newton method. Section IV provides a performance benchmark using the MIT Cheetah 3 dataset. Section V concludes the paper.

II. Preliminaries

A. Physical Consistency of Inertial Parameters

Inertia of a rigid body can be represented by 10 independent parameters that are described with respect to a body-fixed reference frame. The collection of parameters can be written as

$$\phi = \begin{bmatrix} m & h_x & h_y & h_z & I_{xx} & I_{yy} & I_{zz} & I_{yz} & I_{zx} & I_{xy} \end{bmatrix}^T \quad (1)$$

where $m \in \mathbb{R}$ is the mass, $h = \begin{bmatrix} h_x & h_y & h_z \end{bmatrix}^T \in \mathbb{R}^3$ is the first mass moment vector which is the product of mass and centre of mass position vector, and

$$\bar{I} = \begin{bmatrix} I_{xx} & I_{xy} & I_{zx} \\ I_{xy} & I_{yy} & I_{yz} \\ I_{zx} & I_{yz} & I_{zz} \end{bmatrix} \in \mathbb{R}^{3\times 3} \quad (2)$$

is the rotational inertia matrix about the origin of coordinate system.

The inertial parameters of a single rigid body satisfy several inequalities by definition of m, h, and $\bar{I}$ as functionals of a density distribution $\rho(\cdot) : \mathbb{R}^3 \rightarrow \mathbb{R}_+$ (Readers are referred to [9] for the details). Denoting by $\bar{I}_c$ the rotational inertia matrix about the centre of mass and by $J_i = \lambda_i\left(\bar{I}_c\right)$ for $i = 1, 2, 3$ the eigenvalues of $\bar{I}_c$, the full physical consistency conditions for non-degenerate rigid body are given by the following scalar inequalities.

$$\begin{aligned} m &> 0 \\ J_i &> 0 \text{ for } i = 1, 2, 3 \\ J_1 + J_2 &> J_3 \\ J_2 + J_3 &> J_1 \\ J_3 + J_1 &> J_2 \end{aligned} \quad (3)$$

Let us define the notion of pseudo-inertia matrix as

$$L(\phi) \triangleq \begin{bmatrix} \Sigma & h \\ h^T & m \end{bmatrix} \in \mathbb{R}^{4\times 4} \quad (4)$$

where

$$\Sigma \triangleq \frac{1}{2} \operatorname{tr}\left(\bar{I}\right) \mathbb{I}_3 - \bar{I} \quad (5)$$

represents the density-weighted covariance of a rigid body. Note that Σ is linear in ϕ. In [9], [13], the set of constraints in Eq. (3) is shown to be equivalent to a single LMI constraint given by

$$L(\phi) > 0 \tag{6}$$

Equation (6) clearly indicates the non-Euclidean nature of the inertial parameter space.

B. Linear Parameterisation of Rigid Body Dynamics

Consider a n_d degrees-of-freedom system consisting of n_b rigid bodies. The multi-body dynamics can commonly be described as

$$M(q)\dot{\nu} + C(q,\nu)\nu + g_k(q) = F \tag{7}$$

where $q \in \mathcal{Q}$ denotes the configuration lying in the configuration manifold $\mathcal{Q}$, $\nu \in \mathbb{R}^{n_d}$ denotes the generalised velocity, $M(q) \in \mathbb{R}^{n_d \times n_d}$ denotes the mass matrix, $C(q,\nu)\nu \in \mathbb{R}^{n_d}$ denotes the Coriolis force, $g_k(q) \in \mathbb{R}^{n_d}$ denotes the gravitational force, and $F \in \mathbb{R}^{n_d}$ denotes the generalised force. The contributions of joint actuator torques τ, viscous and Coulomb friction forces, external contact wrenches, aerodynamic forces, etc. enter into the dynamics through F. It is well-known that the left-hand-side of Eq. (7) has the linear structure which allows rewriting Eq. (7) as

$$U(q,\nu,\dot{\nu})\Phi = F \tag{8}$$

where $U(q,\nu,\dot{\nu}) \in \mathbb{R}^{n_d \times 10 n_b}$ is the known regressor matrix, and $\Phi \triangleq \begin{bmatrix} \phi^{1T} & \cdots & \phi^{n_b T} \end{bmatrix}^T \in \mathbb{R}^{10 n_b}$ is the collection of inertial parameters of each body. If the generalised force can be described as $F = y - \Lambda(q,\nu,\dot{\nu})\psi$ with some known regressor matrix $\Lambda(q,\nu,\dot{\nu})$, unknown extra parameter vector ψ, and known force element $y \in \mathbb{R}^{n_d}$, then an augmented linear model can be constructed by rearranging Eq. (8) as

$$\underbrace{\begin{bmatrix} U(q,\nu,\dot{\nu}) & \Lambda(q,\nu,\dot{\nu}) \end{bmatrix}}_{\Gamma(q,\nu,\dot{\nu})} \underbrace{\begin{bmatrix} \Phi \\ \psi \end{bmatrix}}_{\theta} = y \tag{9}$$

with the augmented regressor matrix $\Gamma(q,\nu,\dot{\nu})$, and the augmented parameter θ.

C. Bregman Log-Determinant Divergence

Given a strictly convex and differentiable function $\rho(\cdot): S^n \to \mathbb{R}$ that maps the space of real symmetric matrices $S^n \subset \mathbb{R}^{n \times n}$ to $\mathbb{R}$, the Bregman matrix divergence is defined as

$$D_\rho(X,Y) = \rho(X) - \rho(Y) - \langle \nabla\rho(Y), X - Y \rangle \tag{10}$$

where $\langle A, B \rangle = \operatorname{tr}(AB)$. Various Bregman matrix divergences generated by different choices of $\rho(\cdot)$ have been investigated [21].

The log-determinant divergence is a Bregman matrix divergence generated with $\sigma(X) = -\log|X|$ which is defined over the cone of positive definite matrices. It can be expressed as

$$D_\sigma(X,Y) = -\log\frac{|X|}{|Y|} + \operatorname{tr}\left(Y^{-1}X\right) - n \tag{11}$$

The log-determinant divergence has been also known in other names such as Stein's loss [22], Burg matrix divergence, and entropic divergence [11]. The use of log-determinant divergence was studied in the context of low-rank kernel learning [23], [24], information-theoretic metric learning [25], and matrix nearness problem [26], [27].

Given any invertible matrix M, we have $D_\sigma(X,Y) = D_\sigma\left(M^T X M, M^T Y M\right)$. This property suggests that the log-determinant divergence is a coordinate/scale-invariant pseudo-distance metric between two positive definite matrices.

III. Recursive Least Squares with Log-Determinant Divergence Regularisation

A. Regularised Recursive Least Squares with Newton Method

Given N noisy measurements of $\Gamma(q,\nu,\dot{\nu})$ and y, a deterministic approach to estimate the unknown parameter θ in Eq. (9) is to minimise a regularised least squares objective function given by

$$\begin{aligned} J_N(\theta) = & \frac{1}{2}\sum_{j=1}^{N}\left(y_j - \Gamma_j\theta\right)^T W_j \left(y_j - \Gamma_j\theta\right) \\ & - \frac{1}{2}\sum_{j=1}^{N}\left(\theta - \hat{\theta}_j\right)^T G_j \left(\theta - \hat{\theta}_j\right) + \alpha_N R(\theta) \end{aligned} \tag{12}$$

where $y_j \triangleq y(t_j)$, $\Gamma_j \triangleq \Gamma(q(t_j), \nu(t_j), \dot{\nu}(t_j))$, $\hat{\theta}_j \triangleq \hat{\theta}(t_j)$, $W_j = W_j^T > 0$, and $G_j = G_j^T \geq 0$ are the output, the regressor, the estimated parameter, the weighting factor, and the forgetting factor at the j-th time step, respectively; $\alpha_N > 0$ is the regularisation strength at step N and $R(\theta)$ is a twice differentiable convex regulariser.

Let us define $J_0(\theta) = \alpha_0 R(\theta)$ with $\alpha_0 > 0$ and $\hat{\theta}_0 = \underset{\theta}{\arg\min} J_0(\theta)$. Given $\hat{\theta}_{k-1} = \underset{\theta}{\arg\min} J_{k-1}(\theta)$, the recursive update algorithm needs to find the increment Δ_k such that

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \Delta_k = \underset{\theta}{\arg\min} J_k(\theta) \tag{13}$$

The first-order necessary condition for optimality at step $k-1$ gives

$$\begin{aligned} \nabla J_{k-1}\left(\hat{\theta}_{k-1}\right) = & \sum_{j=1}^{k-1}\left(\Gamma_j\hat{\theta}_{k-1} - y_j\right)^T W_j \Gamma_j \\ & - \sum_{j=1}^{k-1}\left(\hat{\theta}_{k-1} - \hat{\theta}_j\right)^T G_j + \alpha_{k-1}\nabla R\left(\hat{\theta}_{k-1}\right) = 0 \end{aligned} \tag{14}$$

By introducing the incremental update given in Eq. (13) and using Eq. (14), the optimality condition at step k requires

$$\begin{aligned} \nabla J_k\left(\hat{\theta}_k\right) & = \sum_{j=1}^{k}\left(\Gamma_j\hat{\theta}_k - y_j\right)^T W_j \Gamma_j \\ & \quad - \sum_{j=1}^{k}\left(\hat{\theta}_k - \hat{\theta}_j\right)^T G_j + \alpha_k \nabla R\left(\hat{\theta}_k\right) \\ & = \left[g_k(\Delta_k)\right]^T + \cancel{\nabla J_{k-1}\left(\hat{\theta}_{k-1}\right)} \\ & = \left[g_k(\Delta_k)\right]^T = 0 \end{aligned} \tag{15}$$

where

$$g_k(\Delta) \triangleq \Gamma_k^T W_k \left(\Gamma_k \hat{\theta}_{k-1} - y_k\right) + \Omega_k \Delta + \left[\alpha_k \nabla R\left(\hat{\theta}_{k-1} + \Delta\right) - \alpha_{k-1} \nabla R\left(\hat{\theta}_{k-1}\right)\right]^T \tag{16}$$

with the information matrix defined as

$$\Omega_k \triangleq \sum_{j=1}^{k} \Gamma_j^T W_j \Gamma_j - \sum_{j=1}^{k-1} G_j \tag{17}$$

The required increment Δ_k can be obtained in closed-form only if $\nabla R(\theta)$ is linear in θ. Otherwise, the algebraic equation given by the last equality in Eq. (15) should be solved for Δ_k numerically.

The convexity and the twice differentiability of $J_k(\theta)$ suggests using Newton-Raphson method to find Δ_k. Let ${}^{l}\Delta_k$ denote the increment for time step k obtained after l Newton-Raphson iterations. Given an initial guess ${}^{0}\Delta_k$ and tolerance ϵ, each iteration proceeds as follows:

1) Compute the Newton step ${}^{l}\delta$ by solving

$$\nabla g_k\left({}^{l-1}\Delta_k\right){}^{l}\delta = -g_k\left({}^{l-1}\Delta_k\right) \tag{18}$$

where

$$\nabla g_k(\Delta) = \Omega_k + \nabla^2 R\left(\hat{\theta}_{k-1} + \Delta\right) \tag{19}$$

2) Compute the Newton decrement

$$\lambda^2 = -\left[g_k\left({}^{l-1}\Delta_k\right)\right]^T {}^{l}\delta \tag{20}$$

3) Stop iteration if $\frac{\lambda^2}{2} \leq \epsilon$
4) (Optional) Find step size γ by backtracking line search
5) Update the solution according to

$${}^{l}\Delta_k = {}^{l-1}\Delta_k + \gamma \cdot {}^{l}\delta \tag{21}$$

The update can be performed in a recursive fashion without storing all pairs of y_j and Γ_j by maintaining only the problem data that are necessary to evaluate $g_k(\Delta)$. One can construct $g_k(\Delta)$ with Ω_k updated as

$$\Omega_k = \Omega_{k-1} - G_{k-1} + \Gamma_k^T W_k \Gamma_k \tag{22}$$

starting from $\Omega_0 = 0$.

B. Log-Determinant Divergence Regularisation

Physically consistent pseudo-inertia matrices are positive definite as shown in Eq. (6). For $i = 1, \cdots, n_b$, the function $D_\sigma\left(L\left(\phi^i\right), L\left(\hat{\phi}_0^i\right)\right)$ represents how close the inertial parameter of the i-th body ϕ^i is to some prior value $\hat{\phi}_0^i$ on the manifold. This motivates the use of log-determinant divergence in [11] as a regulariser in the batch least squares parameter identification framework.

In the same philosophy, this study considers the composite regulariser which takes the log-determinant divergence for the inertial parameters in Φ and the Euclidean distance for extra parameters ψ as

$$R(\theta) = \sum_{i=1}^{n_b} D_\sigma\left(L\left(\phi^i\right), L\left(\hat{\phi}_0^i\right)\right) + \frac{\beta}{2}\left\|\psi - \hat{\psi}_0\right\|_2^2 \tag{23}$$

with constant $\beta > 0$.

Equations (16) and (19) show that the gradient and Hessian of the regulariser $R(\theta)$ are necessary to build the algorithm. The separable structure of Eq. (23) leads to the block structure of the derivatives as

$$\nabla R(\theta) = \left[\nabla_{\phi^1} D_\sigma\left(L\left(\phi^1\right), L\left(\hat{\phi}_0^1\right)\right) \quad \cdots \quad \beta\left(\psi - \hat{\psi}_0\right)^T\right] \tag{24}$$

and

$$\nabla^2 R(\theta) = \begin{bmatrix} \nabla^2_{\phi^1} D_\sigma\left(L\left(\phi^1\right), L\left(\hat{\phi}_0^1\right)\right) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \beta\mathbb{I} \end{bmatrix} \tag{25}$$

Therefore, analytic expressions for the gradient and Hessian of each log-determinant divergence term will suffice the purpose of constructing the required derivatives.

The right superscripts for indexing the bodies are dropped in the followings to avoid cluttered notation. Let ${}^{n}\phi$ denote the n-th component of the inertial parameter vector and $L_0 \triangleq L\left(\hat{\phi}_0\right)$. Since $L(\phi)$ is linear in ϕ, it is obvious that $\frac{\partial L(\phi)}{\partial({}^{n}\phi)} = L(e_n)$ where e_n is the unit vector whose n-th component is 1 and $\frac{\partial^2 L(\phi)}{\partial({}^{m}\phi)\partial({}^{n}\phi)} = 0$. Therefore, using the matrix calculus identity $\frac{\partial \log|U(x)|}{\partial x} = \operatorname{tr}\left[U(x)^{-1}\frac{\partial U(x)}{\partial x}\right]$ yields

$$\frac{\partial D_\sigma(L(\phi), L_0)}{\partial({}^{n}\phi)} = \operatorname{tr}\left[\left({L_0}^{-1} - L(\phi)^{-1}\right)L(e_n)\right] \tag{26}$$

Also, using the linearity of trace operator and another matrix calculus identity $\frac{\partial U(x)^{-1}}{\partial x} = -U(x)^{-1}\frac{\partial U(x)}{\partial x}U(x)^{-1}$ which holds for differentiable and invertible $U(x)$ gives

$$\frac{\partial^2 D_\sigma(L(\phi), L_0)}{\partial({}^{m}\phi)\,\partial({}^{n}\phi)} = \operatorname{tr}\left[L(\phi)^{-1}L(e_m)L(\phi)^{-1}L(e_n)\right] \tag{27}$$

C. Regularisation Strength and Directional Forgetting

The formulation in Eq. 12 provides an enriched framework for modifications through the choice of α_k and G_j. Since α_k determines the relative weighting imposed on the chosen regulariser with respect to the fitting objective, a balanced choice that can reduce performance sensitivity due to the scale difference between the two terms is to set $\alpha_k = \alpha_{sf}\operatorname{tr}\left(\Omega_k\left[\nabla^2 R\left(\hat{\theta}_{k-1}\right)\right]^{-1}\right)$ with constant $\alpha_{sf} > 0$. However, the benefit of adaptive weighting is in a trade-off relationship with the amount of computation. Therefore, using a constant regularisation strength α can be more suitable especially in online identification.

One advantage of online learning is the capability to incorporate a forgetting mechanism with nonzero G_j in Eq. (12). The directional forgetting method of [28] prevents estimator windup by forgetting old information only when it can be updated with new information. This method corresponds to the choice of design parameters given by

$$W_k = \mathbb{I}$$

$$G_{k-1} = \begin{cases} (1-\mu)\,\Omega_{k-1}\Gamma_k^T\left(\Gamma_k\Omega_{k-1}\Gamma_k^T\right)^{-1}\Gamma_k\Omega_{k-1} & \text{if } \left\|\Omega_{k-1}\Gamma_k^T\right\| > \varepsilon \\ 0 & \text{if } \left\|\Omega_{k-1}\Gamma_k^T\right\| \leq \varepsilon \end{cases} \tag{28}$$

with $0 < \mu \leq 1$ and $0 < \varepsilon \ll 1$.

IV. Numerical Experiment

This section benchmarks the proposed algorithm in fixed-base identification of the MIT Cheetah 3 leg assembly. The task is to learn the inertial parameters of each body and the friction coefficients.

A. Model Description

The 3-DOF leg of MIT Cheetah 3 robot which is driven by three proprioceptive actuators is modelled as a system consisting of three links and three rotors, i.e., $n_d = 3$ and $n_b = 6$. The ideal joint space dynamics can be described as

$$\underbrace{\begin{bmatrix} U(q,\dot{q},\ddot{q}) & \Lambda(\dot{q}) \end{bmatrix}}_{\Gamma(q,\dot{q},\ddot{q})} \underbrace{\begin{bmatrix} \Phi^* \\ \psi^* \end{bmatrix}}_{\theta^*} = \tau \tag{29}$$

where $q \in \mathbb{R}^3$ is the joint angles, $U(q,\dot{q},\ddot{q}) \in \mathbb{R}^{3\times 60}$ is the classical regressor, $\Lambda(\dot{q}) = \begin{bmatrix} \mathrm{diag}(\dot{q}) & \mathrm{diag}(\mathrm{sign}(\dot{q})) \end{bmatrix} \in \mathbb{R}^{3\times 6}$ is the friction regressor, $\Phi^* = \begin{bmatrix} \phi_*^{1T} & \cdots & \phi_*^{6T} \end{bmatrix}^T \in \mathbb{R}^{60}$ is the collection of true inertial parameters, $\psi^* \in \mathbb{R}^6$ is the true viscous and Coulomb friction coefficients, and $\tau \in \mathbb{R}^3$ is the joint torques.

B. Experiment Setup

1) Dataset Preparation: The hardware experiment for fixed-base identification of inertial and friction parameters was performed in [9]. This dataset was used in [11], [20] and is made openly available through [12] along with the MATLAB script for computing classical regressor and the CAD data providing initial prior values for the inertial parameters, i.e., $\hat{\Phi}_0 = \hat{\Phi}_{\text{CAD}}$.

This study takes the same dataset which contains a time-series recorded for 28s at the sampling rate of 1kHz. The first 50 and the last 51 datapoints are removed from the dataset since the joint angle values are set to zero and thus the data in those periods are invalid. Therefore, the total number of time points in the dataset spanning the time interval $[0.05\text{s}, 27.95\text{s})$ is $N = 27900$.

2) Experiment Cases: The numerical experiment aims to assess the effects of regulariser and sampling rate on the performance of parameter identification algorithms. The following three methods are compared for verification purposes:

- RLS-l2: RLS with ℓ_2-regularisation for Φ (baseline)
- RLS-ldetdiv: RLS with log-determinant divergence regularisation for Φ (proposed algorithm)
- BLS-ldetdiv: Batch least squares with log-determinant divergence regularisation for Φ (prior works [11], [12])

RLS-l2 considers $R(\theta) = \frac{1}{2}\left\|\Phi - \hat{\Phi}_0\right\|_2^2 + \frac{1}{2}\beta\left\|\psi - \hat{\psi}_0\right\|_2^2$ as the regulariser. By exploiting the Woodbury matrix identity, the update equation for RLS-l2 with $G_j = 0, \forall j$ is given by

$$\begin{aligned} P_k &\triangleq (\Omega_k + W_R)^{-1} \quad \text{where } W_R = \alpha\,\mathrm{diag}(1_{10n_b}, \beta 1_6) \\ &= P_{k-1} - P_{k-1}\Gamma_k^T\left(W_k^{-1} + \Gamma_k P_{k-1}\Gamma_k^T\right)^{-1}\Gamma_k P_{k-1} \qquad (30) \\ \Delta_k &= P_k\Gamma_k^T W_k\left(y_k - \Gamma_k\hat{\theta}_{k-1}\right) \end{aligned}$$

BLS-ldetdiv is implemented in the MATLAB script in [12], and its result is treated as the reference θ^* in this study. The dependency of performance on the sampling rate is tested by downsampling the dataset to specified sampling rate f_s Hz. The size of the downsampled dataset is $N_s = N\frac{f_s}{10^3}$. For a fair comparison, the regularisation strength α for each method is chosen to achieve the same value of $\|\tilde{\tau}(t;t_f)\|_{rms}^{[t_0,t_f]}$ to two decimal places in batch optimisation.

Table I summarises the experiment setup. The `Julia`-based implementation is available via [29].

TABLE I
NUMERICAL EXPERIMENT SETTINGS

Quantity	Value
f_s	$\{10^3, 10^2, 10, 1\}$ [Hz]
$\hat{\theta}_0$	$\begin{bmatrix}\hat{\Phi}_{\text{CAD}}^T & 0^T\end{bmatrix}^T$
W_j	$\frac{1}{N_s n_d}\mathbb{I}_3$
G_j	0
α	50 for RLS-l2, 10^{-1} for RLS-ldetdiv
β	10^{-3}
$\gamma^\dagger$	1
${}^0\Delta_k$	0
ϵ	10^{-20}
$P_0^\ddagger$	$W_R^{-1} = \alpha^{-1}\,\mathrm{diag}(1_{10n_b}, \beta^{-1}1_6)$

† no line search in RLS-ldetdiv, ‡ for RLS-l2

C. Results

The following notations are introduced to clearly describe the results. $\tilde{\theta}(t) = \hat{\theta}(t) - \theta^*$ denotes the parameter error at t. $\hat{\tau}(t;t') = \Gamma(t)\hat{\theta}(t')$ denotes the predicted torque at t considering the parameter estimated at t' and $\tilde{\tau}(t;t') = \hat{\tau}(t;t') - \tau(t)$ denotes the corresponding torque prediction error with respect to the measured torque. The RMS value for a sequence of N_s vector-valued data points $\tilde{\tau}(t;t')$ in the interval $t \in \mathcal{T}$ is defined by

$$\|\tilde{\tau}(t;t')\|_{rms}^{\mathcal{T}} = \sqrt{\frac{1}{N_s}\sum_{j=1}^{N_s}\|\tilde{\tau}(t_j;t')\|_2^2} \tag{31}$$

Figure 1 shows the time histories of the geometric distance for $\hat{\Phi}$ given by $D_\Phi \triangleq \sum_{i=1}^{n_b} D_\sigma\left(L\left(\hat{\phi}^i\right), L\left(\phi_*^i\right)\right)$, the Euclidean norm for $\hat{\psi}$ given by $D_\psi \triangleq \left\|\left(\hat{\psi} - \psi^*\right)/\psi^*\right\|_2$ with elementwise division denoted by $/$ for normalisation, $\left\|\tilde{\theta}(t)\right\|_2$, and $\|\tilde{\tau}(t';t)\|_{rms}^{[t,t_f]}$ for the case of $f_s = 10^3$ Hz. D_Φ cannot be computed for the estimate provided by RLS-l2 due to physical inconsistency. Unlike RLS-l2, $\hat{\theta}(t)$ converges to θ^* with RLS-ldetdiv. The RMS value shown is for the error in the output predicted using current parameter estimate for unseen future state. The result indicates better generalisation performance of RLS-ldetdiv in the low-data regime as compared to RLS-l2.

Figure 2 shows the time histories of the measured torque $\tau(t)$ and the predicted torque $\hat{\tau}(t;t)$ for each method being compared. Note that BLS-ldetdiv is only capable of offline prediction, thus the distinction of $\hat{\tau}(t;t')$ depending on t' does not apply to BLS-ldetdiv. The prediction of online identification methods show deviation from the result of BLS-ldetdiv during the first few seconds. Both RLS-l2 and RLS-ldetdiv managed to average out the effect of noise in the measurements as the stream of data accumulates. However, only the predicted torque of RLS-ldetdiv converges to that of BLS-ldetdiv since $\hat{\theta}(t) \to \theta^*$ in RLS-ldetdiv.

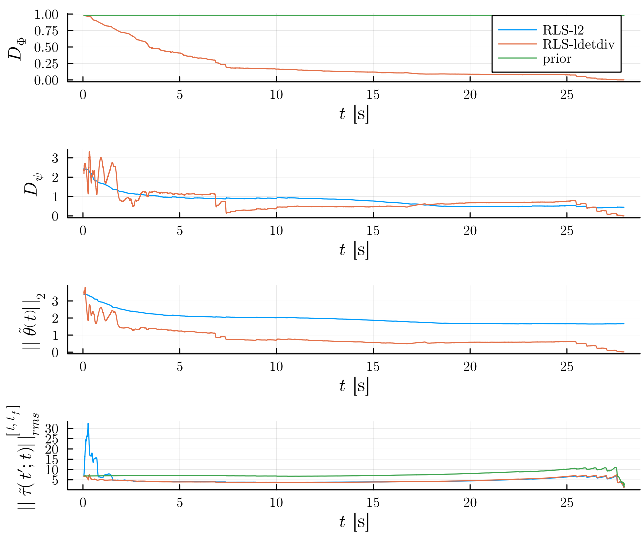

Fig. 1. Parameter and Torque Prediction Errors ($f_s = 10^3$ Hz)

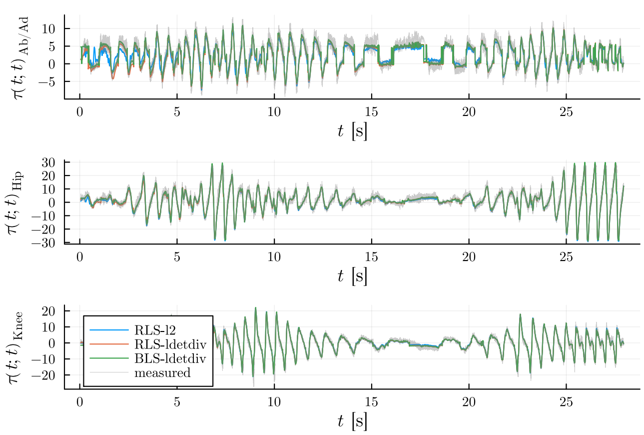

Fig. 2. Measured and Predicted Joint Torques ($f_s = 10^3$ Hz)

Table II shows the benchmark result for a range of f_s with the indicator for full physical consistency of $\hat{\Phi}(t_f)$ given by $\mathcal{I}_{cons} \triangleq \prod_{i=1}^{n_b} \mathsf{isposdef}\left(L\left(\hat{\phi}^i(t_f)\right)\right)$, the indicator for positive definiteness of the estimated mass matrix given by $\mathcal{I}_M \triangleq \prod_{j=1}^{N_s} \mathsf{isposdef}\left(M\left(q(t_j);\hat{\Phi}(t_j)\right)\right)$, $\left\|\tilde{\theta}(t_f)\right\|_2$, $\|\tilde{\tau}(t;t)\|_{rms}^{[t_0,t_f]}$, and $\|\tilde{\tau}(t;t_f)\|_{rms}^{[t_0,t_f]}$. The RMS value of torque prediction error is comparable as expected from the choice of α for each method. However, only RLS-ldetdiv provides fully physically consistent estimates for the inertial parameters regardless of f_s. Moreover, the mass matrix estimated by RLS-ldetdiv is always positive definite which complies with the essential requirement in many robot control schemes. The accuracy of RLS-ldetdiv with respect to θ^* obtained by BLS-ldetdiv using entire dataset is improved with more data points, whereas increasing f_s does not lead to effective reduction in $\left\|\tilde{\theta}(t_f)\right\|_2$ for RLS-l2.

TABLE II
Sampling Rate Dependency of Performance Metrics

		f_s [Hz]			
		10^0	10^1	10^2	10^3
$\mathcal{I}_{cons}$	RLS-l2	0	0	0	0
	RLS-ldetdiv	1	1	1	1
$\mathcal{I}_M$	RLS-l2	0	0	0	0
	RLS-ldetdiv	1	1	1	1
$\left\|\tilde{\theta}(t_f)\right\|_2$	RLS-l2	1.956	1.625	1.659	1.663
	RLS-ldetdiv	1.158	0.347	0.067	0.014
$\|\tilde{\tau}(t;t)\|_{rms}^{[t_0,t_f]}$	RLS-l2	3.004	3.554	3.585	3.619
	RLS-ldetdiv	3.279	3.554	3.507	3.532
$\|\tilde{\tau}(t;t_f)\|_{rms}^{[t_0,t_f]}$	RLS-l2	4.030	3.577	3.397	3.407
	RLS-ldetdiv	4.130	3.593	3.395	3.404

Computation time per each time step is assessed for RLS-ldetdiv on Macbook Pro 15-inch 2017 with 2.8GHz Intel Core i7 CPU and 16GB 2133MHz LPDDR3 RAM running Julia 1.9.3. At step k, $\left[L\left(\hat{\phi}_{k-1}^i\right)\right]^{-1} L(e_n)$ for $i = 1, \cdots, n_b$ and $n = 1, \cdots, 10$ are computed before starting the Newton-Raphson iterations to reduce the number of matrix operations. The increment Δ_k in RLS-ldetdiv is found in no more than two Newton-Raphson iterations at each time step even for the tight stopping criterion specified by ϵ. The mean and the standard deviation of the per-step computation time for the cases ended in one Newton-Raphson iteration are 1.623ms and 0.340ms, respectively, and those for two Newton-Raphson iterations are 2.199ms and 0.421ms, respectively. Implementation using a low-level language such as C or C++ is expected to provide further improvements in runtime efficiency.

V. Conclusion

An online parameter estimator is developed based on linear regression with the recursive regularised least squares approach. The log-determinant divergence defined over the cone of positive definite pseudo-inertia matrices is used as the regulariser to keep the full physical consistency of estimated inertial parameters. The analytic expressions for the gradient and Hessian of the regulariser help efficient pointwise computation of the increment using the Newton-Raphson method. Recursive update of the information matrix allows online identification without storing every input-output pair in the memory and incorporation of optional forgetting factors through problem modification. Numerical experiment using the MIT Cheetah 3 dataset confirmed that the recursive implementation converges to the same result as the batch optimisation with at most two Newton-Raphson iterations at each time. The proposed algorithm provides fully physically consistent sequence of inertial parameters and positive definite mass matrix regardless of the sampling rate. The benchmark results support the validity and practicality of the proposed algorithm.

REFERENCES

[1] Z. R. Manchester and M. A. Peck, "Recursive Inertia Estimation with Semidefinite Programming," in *AIAA Guidance, Navigation, and Control Conference, SciTech Forum*, Grapevine, TX, USA, January 2017.

[2] J. Ahmed, V. T. Coppola, and D. S. Bernstein, "Adaptive Asymptotic Tracking of Spacecraft Attitude Motion with Inertia Matrix Identification," *Journal of Guidance, Control, and Dynamics*, vol. 21, no. 5, pp. 684–691, 1998.

[3] M. C. Norman, M. A. Peck, and D. J. O'shaughnessy, "In-Orbit Estimation of Inertia and Momentum-Actuator Alignment Parameters," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 6, pp. 1798–1814, 2011.

[4] H. Yoon, K. M. Riesing, and K. Cahoy, "Kalman Filtering for Attitude and Parameter Estimation of Nanosatellites Without Gyroscopes," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 9, pp. 2272–2288, 2017.

[5] P. Nuthi and K. Subbarao, "Computational Adaptive Optimal Control of Spacecraft Attitude Dynamics with Inertia-Matrix Identification," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 5, pp. 1258–1262, 2017.

[6] E. A. Morelli, "Determining Aircraft Moments of Inertia from Flight Test Data," *Journal of Guidance, Control, and Dynamics*, vol. 45, no. 1, pp. 4–14, 2022.

[7] J. Geng and J. W. Langelaan, "Estimation of Inertial Properties for a Multilift Slung Load," *Journal of Guidance, Control, and Dynamics*, vol. 44, no. 2, pp. 220–237, 2021.

[8] S. Traversaro, S. Brossette, A. Escande, and F. Nori, "Identification of Fully Physical Consistent Inertial Parameters Using Optimization on Manifolds," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Daejeon, South Korea, October 2016.

[9] P. M. Wensing, S. Kim, and J.-J. E. Slotine, "Linear Matrix Inequalities for Physically Consistent Inertial Parameter Identification: A Statistical Perspective on the Mass Distribution," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 60–67, 2018.

[10] T. Lee and F. C. Park, "A Geometric Algorithm for Robust Multibody Inertial Parameter Identification," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2455–2462, 2018.

[11] T. Lee, P. M. Wensing, and F. C. Park, "Geometric Robot Dynamic Identification: A Convex Programming Approach," *IEEE Transactions on Robotics*, vol. 36, no. 2, pp. 348–365, 2020.

[12] P. M. Wensing, "inertia_identification_minimal_examples," GitHub, 2022.

[13] C. D. Sousa and R. Cortesão, "Physical Feasibility of Robot Base Inertial Parameter Identification: A Linear Matrix Inequality Approach," *The International Journal of Robotics Research*, vol. 33, no. 6, pp. 931–944, 2014.

[14] T. Lee, J. Kwon, and F. C. Park, "A Natural Adaptive Control Law for Robot Manipulators," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Madrid, Spain, October 2018.

[15] C. D. Sousa and R. Cortesão, "Inertia Tensor Properties in Robot Dynamics Identification: A Linear Matrix Inequality Approach," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 1, pp. 406–411, 2019.

[16] T. Lee, J. Kwon, P. M. Wensing, and F. C. Park, "Robot Model Identification and Learning: A Modern Perspective," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 7, no. 1, 2024.

[17] Q. Leboutet, J. Roux, A. Janot, J. R. Guadarrama-Olvera, and G. Cheng, "Inertial Parameter Identification in Robotics: A Survey," *Applied Sciences*, vol. 11, no. 9, pp. 1–41, 2021.

[18] A. Janot and P. M. Wensing, "Sequential Semidefinite Optimization for Physically and Statistically Consistent Robot Identification," *Control Engineering Practice*, vol. 107, no. 104699, pp. 1–15, 2021.

[19] T. Lee, "Geometric Methods for Dynamic Model-Based Identification and Control of Multibody Systems," Ph.D. dissertation, Department of Mechanical and Aerospace Engineering, Seoul National University, 2019. [Online]. Available: https://hdl.handle.net/10371/161901

[20] C. Rucker and P. M. Wensing, "Smooth Parameterization of Rigid-Body Inertia," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2771–2778, 2022.

[21] A. Cichocki, S. Cruces, and S.-i. Amari, "Log-Determinant Divergences Revisited: Alpha-Beta and Gamma Log-Det Divergences," *Entropy*, vol. 17, no. 5, pp. 2988–3034, 2015.

[22] W. James and C. Stein, "Estimation with Quadratic Loss," in *4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, Berkeley, CA, USA, June-July 1960, p. 361–379.

[23] B. Kulis, M. A. Sustik, and I. S. Dhillon, "Learning Low-Rank Kernel Matrices," in *23rd International Conference on Machine Learning*, Pittsburgh, PA, USA, June 2006, p. 505–512.

[24] ——, "Low-Rank Kernel Learning with Bregman Matrix Divergences," *Journal of Machine Learning Research*, vol. 10, no. 13, pp. 341–376, 2009. [Online]. Available: http://jmlr.org/papers/v10/kulis09a.html

[25] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-Theoretic Metric Learning," in *24th International Conference on Machine Learning*, Corvallis, OR, USA, June 2007, pp. 209–216.

[26] I. S. Dhillon and J. A. Tropp, "Matrix Nearness Problems with Bregman Divergences," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 4, pp. 1120–1146, 2008.

[27] I. S. Dhillon, "The Log-Determinant Divergence and its Applications," in *Householder Symposium XVII*, Zeuthen, Germany, June 2008.

[28] L. Cao and H. Schwartz, "A Directional Forgetting Algorithm Based on the Decomposition of the Information Matrix," *Automatica*, vol. 36, no. 11, pp. 1725–1731, 2000.

[29] N. Cho, "EntDiv_RLS," GitHub, 2023.