Discrete-Time Elasto-Plastic Friction Estimation

Vincent Hayward, Fellow, IEEE, Brian S. R. Armstrong, Senior Member, IEEE, Friedhelm Altpeter, Member, IEEE, and Pierre E. Dupont, Senior Member, IEEE

Abstract—For control applications involving small displacements and velocities, friction modeling and compensation can be very important, especially around velocity reversal. We previously described single-state friction models that are based on elasto-plastic presliding, something that reduces drift while preserving the favorable properties of existing models (e.g., dissipativity) and that provide a comparable match to experimental data. In this paper, for this class of models, discrete estimation for friction force compensation is derived. The estimator uses only position and velocity (not force) measurements and integrates over space rather than time, yielding a discrete-time implementation that is robust to issues of sample size and sensor noise, reliably renders static friction and is computationally efficient for real-time implementation. Convergence during steady sliding and dissipativity are established for the discrete-time formulation.

Index Terms—Discrete-time control, friction compensation, friction estimation, friction modeling, mechanical systems, presliding.

I. INTRODUCTION

FOR CONTROL, friction modeling can be complex, particularly to account for the details of friction around velocity reversals, at very low velocity or for very small motions. Over the past 40 years, it has become clear that dynamic friction phenomena play a role under these conditions. The literature addressing dynamic friction and control is large and growing by several tens of articles each year. No attempt is made to survey the literature here, but see [1]–[5].

Some applications of dynamic friction models require real-time implementation. An example is rendering friction as part of simulating object manipulation in a virtual reality, which is needed to drive a haptic man-machine interface [6], [7]. Modeling static friction is particularly important for this application, since without it objects can drift about in the virtual reality. Thus, a suitable friction model for haptic rendering must both reliably render static friction and be suitable for real-time implementation.

Prior to the introduction of the state variable models, simulating velocity reversal and static friction often involved discontinuities in the differential equation and even a change of model structure to reflect static friction. For example, the Karnopp model changes from one dynamic equation to another to model static friction [8]. Similarly, the reset integrator model described in [4] has one structure for static friction and another with one fewer states for sliding friction. Both of these models have been applied to haptic rendering [9], [10].

Friciton models that change structure in the transition from static to sliding friction have inherent disadvantages. Sensor noise can be sufficient to force a transition between one structure and another. Additionally, when a transition from sliding to sticking is detected, it can be necessary to back up the simulation to the point of the transition and restart it in the new structure. This complexity is always undesirable and may be unsuitable for real-time implementation.

The Dahl-type friction models [1], particularly the LuGre model [5], overcome the need to have two structures in the friction model. A single state variable qualitatively models the state of elastic deformation in the contact, as illustrated in Fig. 1. The state variable $z(t)$ is a proxy for the deflection of all asperity contacts between the two sliding surfaces. With the LuGre model, the Stribeck effect and frictional lag can be modeled with a single state equation and no discontinuities [5]. This model has been applied in numerous studies [11]–[13].

The Dahl-type friction models, however, have been studied in the continuous time and discrete-time implementations face a challenge because of the nonlinear and stiff nature of the dynamics of the elastic state $z(t)$ [14]. In offline simulations, a sophisticated numerical integrator with variable step size can be used. But these methods may be impractical for real-time implementation with limited computing power, such as a haptic interface driven by a microcontroller [6]. Additionally, both the Dahl and LuGre models exhibit drift rather than static friction [15].
To be suitable for applications requiring real-time implementation and where representing static friction is important two improvements are required. The state equation must be modified to model static friction, leading to the elasto-plastic (EP) model [15], and an efficient, discrete-time formulation is needed for implementation in a microcontroller. The remainder of the paper is organized as follows. Static friction in state-variable models is discussed in Section II, and the discrete-time EP model is presented in Section III. Several desirable properties of the model are demonstrated in Section IV, and robustness to implementation issues is addressed in Section V. Finally, we conclude in Section VI.

II. STATIC FRICTION IN DAHL-TYPE STATE-VARIABLE FRICTION MODELS

Referring to Fig. 1, where $z(t)$ represents the externally measurable rigid body displacement of the mass; $z(t)$ represents the deflection of the proxy asperity or the elastic component of the displacement, and $w(t)$ represents the sliding or plastic component. The total displacement $x(t)$ is the sum of the elastic and plastic components, $z(t)$ and $w(t)$, respectively

$$x(t) = z(t) + w(t). \quad (1)$$

The “elasto-plastic” model studied by Prandtl to represent the behavior of solids under stress is applied here to represent the regimes of motion possible in the Dahl-type friction models [17]. Also seen in Fig. 1, term $f_d(t)$ is the externally applied force on the body and $f_f(t)$ is the friction force.

Several Dahl-type friction models have been presented with different forms for the evolution of the internal state. In the original Dahl model, the state is governed by [1]

$$\dot{z}(t) = \left(1 - \frac{z(t)}{z_C}\right) \dot{x}(t), \quad (2)$$

in the LuGre model, the state is governed by [5]

$$\dot{z}(t) = \left(1 - \frac{z(t)}{z_{ss}(\dot{x}(t))}\right) \dot{x}(t) \quad (3)$$

and in the EP model, it is governed by [15]

$$\dot{z}(t) = \left(1 - \alpha(z, \dot{x}) \frac{z}{z_{ss}(\dot{x})}\right) \dot{x}(t) \quad (4)$$

where $z_C = f_C/\sigma_0$ is the asperity deflection corresponding to Coulomb friction and $z_{ss}(\dot{x}(t))$ is the value of $z(t)$ corresponding to steady-state sliding with velocity $\dot{x}(t)$. With each of these forms is associated an output equation of the form

$$f_f(t) = \sigma_0 z(t) + \sigma_1 \dot{z}(t) + \sigma_2 \ddot{z}(t) \quad (5)$$

with

$$\sigma_0 > 0, \quad \sigma_1, \sigma_2 \geq 0$$

where $f_f(t)$ is the instantaneous friction force determined from the model, parameter $\sigma_0$ models the stiffness of the frictional contact; $\sigma_1$ can be tuned to damp oscillations of state $z(t)$; $\sigma_2$ is the viscous friction parameter. Dahl incorporated exponent $i$ to shape the force-displacement curve to measured experimental data and did not incorporate any term $\sigma_1$ [1]. Empirical results suggest that $i = 1$ is adequate.

In lubricated contacts, the friction-velocity curve, which plots steady state friction force, $f_{ss}$, versus velocity, $\dot{x}(t)$, shows a smooth rather than abrupt transition from static to sliding friction, such as is illustrated in Fig. 2. Depending on the detail with which friction at low velocities is to be modeled, several forms are commonly used for $f_{ss}(\dot{x}(t))$. These forms are

$$f_{ss}(\dot{x}(t)) = f_C \text{sgn}(\dot{x}(t)) \quad (6)$$

$$f_{ss}(\dot{x}(t)) = \left[(f_{\text{max}} - f_C)e^{-\left(\dot{x}(t)/v_{\text{ss}}\right)^2} + f_C\right] \text{sgn}(\dot{x}(t)) \quad (7)$$

$$f_{ss}(\dot{x}(t)) = \left[(f_{\text{max}} - f_C)\frac{1}{1 + \left(\frac{\dot{x}(t)}{v_{\text{ss}}}\right)^2} + f_C\right] \text{sgn}(\dot{x}(t)) \quad (8)$$

where (6) provides a simple Coulomb friction model and was used with (2) by Dahl [1]; (7) is used in the LuGre model [5]; and (8) provides a model that is practically equivalent to (7) and is more efficient to compute [2]. In (6)–(8), $f_C$ is the Coulomb friction level; $f_{\text{max}}$ is the maximum or static friction level, and $v_{\text{ss}}$ is the characteristic velocity of the Stribeck friction. The term $z_{ss}(\dot{x}(t))$ in the Dahl-type models reflects the steady-state deflection of the hypothetical asperity and is given by

$$z_{ss}(\dot{x}(t)) = \begin{cases} \frac{f_{\text{ss}}(\dot{x}(t))}{\sigma_0}, & |\dot{x}| > 0 \\ \frac{f_{\text{max}}}{\sigma_0}, & \dot{x} = 0 \end{cases} \quad (9)$$

where the case statement ensures that $z_{ss}(\dot{x}(t))$ is defined when $\dot{x} = 0$. With each of (2)–(4), $z(t) \rightarrow z_{ss}(\dot{x}(t))$ during constant velocity sliding.

A key advantage of the Dahl-type models is that they render friction that varies in a smooth way as velocity goes through zero, avoiding discontinuities in the friction force and changes to the model structure (e.g., number of state variables) required by the Coulomb- or Karnopp-type models at zero velocity. The Dahl-type models incorporate $z_{ss}(\dot{x}(t))$, which is seen in (6)–(9) to be discontinuous at $\dot{x}(t) = 0$. The distinction between the Dahl-type models and the discontinuous models is that for the Dahl-type models $\dot{z}(t)$, $z(t)$ and $f_f(t)$ are all continuous functions of time. This is seen by considering (2)–(4), where $\dot{x}(t)$ multiplies the term with $z_{ss}(\dot{x})$. The right-hand side of the differential equation is equal to zero when $z_{ss}(\dot{x}(t))$ encounters its discontinuity.
While simplified, the elasticity illustrated in Fig. 1 and associated with \( z(t) \) in (1)–(2) is physical. Elastic behavior in frictional contacts is directly observed, for example, in pointing systems and high-precision machine tools [1], [16]. It is the discontinuous nature of Coulomb-type friction models which is non-physical.

The everyday notion of static friction assumes that \( \dot{z} \equiv 0 \) for \( f_a(t) < f_s \). However, when contact elasticity is considered the notion of static friction must be augmented to accommodate small elastic motions that can arise even when there is no plastic motion. Informally, a contact in static friction may exhibit elastic motions, but there should be no cumulative motion when \( f_a(t) < f_s \). Formalizing this idea, a state-variable friction model possesses a static friction or stiction phase if there exists a breakaway force \( f_{ba} \) such that, for any friction force \( f_f(t) \) that satisfies

\[
|f_f(t)| < f_{ba}, \quad t \in [t_0, t_f]
\]

all motions are bounded in the range

\[
|x(t) - x(t_0)| \leq 2z_{\text{max}}, \quad t \in [t_0, t_f]
\]

where \( t_f > t_0 \) and \( z_{\text{max}} \) is the maximum elastic deformation of the frictional contact.

Following Prandtl’s description of elastoplastic deformation, motion in a friction contact can be characterized by three regimes: purely elastic displacement, mixed elastic and plastic displacement and purely plastic displacement [17]. Using (1), the three regimes are quantitatively determined by

\[
\begin{aligned}
\dot{x} &= \dot{z} \\
\dot{v} &= 0
\end{aligned}
\]  

purely elastic displacement,

\[
\begin{aligned}
\dot{x} &= \dot{z} + \dot{v} \\
\dot{v} &= \dot{w}
\end{aligned}
\]  

mixed elastic and plastic displacement,

\[
\begin{aligned}
\dot{x} &= \dot{v} \\
\dot{v} &= 0
\end{aligned}
\]  

sliding.

For a state-variable friction model to exhibit static friction, it is sufficient that

\[
\frac{\partial w}{\partial t} = 0, \quad \text{for all } |f_a(t)| < f_{ba}
\]

since otherwise there generally exists a force trajectory \( f_f(t) \), \( t_1 \leq t \leq t_2 \), such that \( z(t_1) = z(t_2) \) but \( w(t_1) \neq w(t_2) \). Writing \( \delta w = w(t_2) - w(t_1) \), by simply repeating this trajectory \( N \) times, with \( N \) sufficiently large, the \( \delta w \) accumulate so that \( |x(t) - x(t_0)| \approx |N \delta w| > 2z_{\text{max}} \). Two examples involving oscillatory motion show the possible behaviors of the Dahl-type friction models. In the examples below, units are given for linear motion, but could be expressed for rotational motion with suitable change of dimension.

The first example is a simulation with parameters given in Table I and results shown in Fig. 3. The simulation is done with (3) without an elastic regime, and with (4) having an elastic regime. At \( t = 0 \) and \( t = 8 \), \( f_a = f_0 = 0 \). In the interval \( 0 < t < 6 \) an oscillatory signal is applied. For the model without an elastic regime, \( \dot{w}(t) \neq 0 \) and \( x(t) \neq x(t_0) \). The example shows that for this model, as the frequency of the oscillatory input increases, the rate of drift increases.

The second example is taken from experimental data from an instrumented electrical discharge machining (EDM) system [16]. The machine is instrumented to measure friction during low-velocity slideway motions [16]. Measured friction data are seen in Fig. 4(c). The position sensor data recorded in the experiment (not shown, see [16]) include a noise component. Friction
III. DISCRETE-TIME ELASTO-PLASTIC FRICTION
MODEL IMPLEMENTATION

Introduction of \( \alpha(z, \dot{x}) \) meets the first goal of introducing static friction into the state variable model. The second goal is an efficient discrete time formulation. For discrete-time analysis, define a sequence of sample times \( \{ t_k \} \), not necessarily uniformly sampled in time. For example, sampling can be driven by position sensor interrupts in some applications. Define \( \{ x_k \} \) as the sequence of rigid body displacements at times \( \{ t_k \} \) and \( \{ \dot{x}_k \} \) as the sequence of measurements or estimates of \( \dot{x}(t_k) \), and \( \{ \ddot{x}_k \} \) as a sequence of estimates of \( \ddot{x}(t_k) \). The dynamics of the state variable model can be written in time-free form by relating differential quantities of displacement, that is

\[
\frac{d z}{d x} = 1 - \alpha(z, \dot{x}) \frac{z}{z_{ss}(\dot{x})},
\]

Next, integrate this expression over a displacement \( x_{k+1} - x_k \)

\[
z_k - z_{k-1} = (x_k - x_{k-1}) - \int_{x_{k-1}}^{x_k} \alpha(z, \dot{x}) \frac{z}{z_{ss}(\dot{x})} d x
\]

\[
= (x_k - x_{k-1}) - (w_k - w_{k-1}).
\]

For a real-time implementation, the most straightforward integration of (11) is by explicit Euler integration. As is shown below, when the step size \( \Delta \tau_k = \tau_k - \tau_{k-1} \) is small relative to \( z_{ss} \), boundedness and convergence of the discrete time model can be shown. This result may not be practical for real-time implementation, however, because \( z_{ss} \) may be quite small, and in the usual case of periodic sampling, a bound on \( \Delta \tau_k \) may translate into an impractically small bound on velocity. To ensure that the model deteriorates gracefully as \( \Delta \tau_k \) increases, we propose the simple expedient of saturating the sampled sequence \( \dot{z}_k \) at its maximum value of \( z_{ss}(\dot{v}_k) \). This method is an extension to that described in [6] to the dynamic case. Because \( z_{ss}(\dot{v}_k) \) can change between sample \( t_{k-1} \) and \( t_k \), it is necessary to saturate \( \dot{z}_k \) twice, once as it is computed, and again as it is used for the next update. Incorporating the two saturation operations gives the update law for the discrete-time, elasto-plastic friction model, as expressed by

\[
y_k = \begin{cases} 
1 - \alpha(z_{k-1}, \dot{v}_k) \frac{z_{k-1}}{z_{ss}(\dot{v}_k)}, & \text{if } \frac{z_{k-1}}{z_{ss}(\dot{v}_k)} < 1 \\
0, & \text{otherwise}
\end{cases}
\]

\[
\dot{z}_k = \begin{cases} 
\dot{z}_{k-1} + y_k, & \text{if } \frac{z_{k-1}}{z_{ss}(\dot{v}_k)} < 1 \\
z_{ss}(\dot{v}_k), & \text{otherwise}
\end{cases}
\]

where \( y_k \) is the unsaturated estimate of the change in \( \dot{z} \). Equations (12) and (13), together with (8) and (10) give the discrete-time, elasto-plastic friction model. A saturation mechanism of this type has been incorporated by the third author into high-precision machine tool control, with many units deployed around the world. Note, however, that using this estimated signal for friction compensation in closed loop may lead to chattering. Capturing the case statements of (10), (12) and (13) with a

estimated by a LuGre model is seen in Fig. 4(a). With low-velocity motion and in the presence of sensor noise, friction is significantly underestimated. This arises because the model lacks a purely elastic regime and, in a fashion similar to the response to a small oscillation in Fig. 3, the noise in the data causes \( z(t) \) to relax, which in turn results in an underestimation of the friction. Fig. 4(b) presents friction estimated by the EP model, which incorporates an elastic displacement regime. In the elastic regime, noise-induced drift is eliminated and friction is reliably estimated even for low-velocity motions.

A third example is haptic rendering, which provided the initial motivation for the developments reported here [6]. In haptic rendering, \( x(t) \) is sensed as the user manipulates the haptic input, and \( f_j(t) \) is computed and applied as force feedback through the haptic device (see [7, Fig. 3, right panel]). Because small tremors are natural to the human hand, the signal \( x(t) \) inevitably includes an oscillatory component similar to that seen in Fig. 4(a). When a Dahl-type friction model is used, objects in the simulated reality drift perceptibly. An EP model overcomes the challenge by incorporating a purely elastic regime.

Relating (1) and (4), the rate of plastic displacement in the EP model is given by

\[
\dot{w}(t) = \alpha(z, \dot{x}) \frac{z(t)}{z_{ss}(t)} \dot{x}(t)
\]

with \( \dot{w}(t) = 0 \) for elastic displacement. By setting conditions for \( \alpha(z, \dot{x}) \) to 0, the conditions for static friction can be controlled. There are many possible choices for \( \alpha(z, \dot{x}) \). A form like that seen in Fig. 5 gives a region \( \{ z(t) : -z_{in} \leq z(t) \leq z_{in} \} \) on which \( \dot{w}(t) = 0 \). When \( \sigma_0 > 0 \) and \( \sigma_1, \sigma_2 \geq 0 \), it is straightforward to show that \( |f_j(t)| < f_j(z) \) for \( \dot{x} < z_{in} \) on the same interval, and thus a purely elastic regime is provided and static friction is modeled. One suitable choice for \( \alpha(z, \dot{x}) \) is the piecewise continuous function

\[
\alpha(z, \dot{x}) = \begin{cases} 
0, & \text{if } \text{sgn}(z) \neq \text{sgn}(\dot{x}), \\
\sigma_0, & \text{if } \text{sgn}(z) = \text{sgn}(\dot{x}), \quad |z| \leq z_{in} \\
\sin \left( \frac{|z|}{z_{ss}(\dot{x})} \right) \left( 1 - \frac{z_{ss}(\dot{x})}{z_{in}} \right) + \frac{1}{2}, & \text{if } 2z_{in} < |z| < z_{ss}(\dot{x}) \\
1, & \text{if } |z| \geq z_{ss}(\dot{x})
\end{cases}
\]

where \( \sigma_0 \) and \( z_{in} \) are parameters, and \( z_{ss}(\dot{x}) \) is a function which models Strubeck friction.

Fig. 5. Candidate curve to model transition from plastic to elastic behavior. Under the boundary deflection, \( z_{in}, \alpha(z) = 0 \), and the model behaves elastically. Above \( z_{in}, \) the contact is sliding and the steady-state deflection follows the dynamics of Strubeck friction as in Fig. 2. A member of the family of resulting curves is symbolized by a dashed line. The critical case when \( \dot{x} = \tau_{in} \) is shown by a solid line.
model for Stiebeck friction, the discrete-time EP model can be expressed in pseudo code as shown in Algorithm 1.

```plaintext
**** Pseudo Code, discret-time EP model ****/
/* Model Stiebeck curve */
float fss(vbarK)
    return((fmax-fc)/(1+(vbarK/VS)²)+fC)*sign(vbarK);
}
/* Determine alpha(z, xdot) */
float alpha(zhatK, vbarK)
    float alpha, zss = fss(vbarK)/sigma:0;
    if(zhatK>vbarK <= 0) alpha = 0;
    elseif(abs(zhatK) <= zba) alpha = 0;
    elseif(abs(zhatK) >= zss) alpha = 1;
    else alpha = 0.5*sin(pi*abs(zhatK-(zss + zba)/zss-zba)) + 0.5;
    return alpha;
}
/* Determine zhat(k), the next zhat */
float Update_zhat(zhatK, Delta_xbarK, vbarK)
    float zk, zhatK;
    float zss = fss(vbarK)/sigma:0;
    if (zhatK/zss < 1)
        yk =
            (1 - alpha(zhatK,vbarK)*(zhatK/zss)) * Delta_xbarK;
    else
        yk = 0;
    if (abs(zhatK + yk)/zss < 1) zhatK = zhatK + yk;
    else zhatK = zss;
    return zhatK;
```

Algorithm 1. The cases of function \( \alpha(\cdot) \) assure static friction for \( |z(t)| < z_{ba} \), while the cases of \( \text{Update}_z\hat{\text{hat}} \), assure the boundedness of \( z_k \) and convergence during steady sliding. Terms \( f_{\text{max}}, f_C, V_S \) and \( Z_BA \) are model parameters; argument \( \Delta x_{\text{bar}} \) is \( \Delta r_k \) and \( V_{\text{bar}} \) is \( \eta_k \); and argument term \( \text{zhatK} \) is the previous value of the state, \( z_{k-1} \).

The EP model requires the parameters to be arranged so that

\[
\sigma_0 > 0, \quad V_S > 0, \quad \sigma_1 \geq 0, \quad \sigma_2 \geq 0, \\
0 < z_{sa} < z_C \leq [z_{\text{ss}} (\hat{z})] \leq z_{\text{max}}
\]

and with signals \( \Delta r_k \) and \( \eta_k \) prepared so that \( \text{sgn}(\Delta r_k) = \text{sgn}(\eta_k) \). Analogous to (5), friction is given by

\[
\hat{f}_f(t_k) = \sigma_0 \dot{z}_k + \sigma_1 \left( \frac{\Delta z_k}{\Delta t_k} \right) + \sigma_2 \eta_k
\]

where \( \Delta z_k = \dot{z}_k - \dot{z}_{k-1} \) and \( \Delta t_k = t_k - t_{k-1} \). Convergence and the other properties of the EP model are demonstrated in Section IV. Performance of the model is illustrated by an example in Section V.

IV. DEMONSTRATION OF THE PROPERTIES OF THE DISCRETE-TIME MODEL

Properties that are important for state-variable friction models include: bounded-input bounded-output stability of the friction estimator, during steady sliding with \( v(t) = \bar{v} \) the estimator state converges to its steady-state value, during sliding friction opposes slip, and the model is dissipative for all \( \Delta r_k \neq 0 \). These properties are formally demonstrated in theorems 1–4 in the following. To ensure these properties, and indeed numerical stability, the saturation operations of (12) and (13) are required for the discrete time model. The saturation operations of (12) and (13) create three cases for the friction estimation update and several of the proofs must treat the cases individually. The three cases of (12) and (13) are first described and then the properties of theorems 1–4 are established.

A. Description of the Three Cases of the Friction Estimation Update

The friction estimator, (12) and (13), includes two case statements, giving a total of four combinations. Of these, it is not possible to take the lower branch of (12) and the upper branch of (13), and so only three cases are possible.

Case 1) Corresponds to the lower branch in each of (12) and (13), and arises when \( \hat{z}_{k-1}/z_{ss} (\hat{\eta}_k) \geq 1 \). This case can be triggered if the system is accelerating through Stiebeck friction and \( f_{ss} (\hat{\eta}_k)/f_{ss} (\hat{\eta}_{k-1}) < 1 \), and gives \( \hat{z}_k = z_{ss} (\hat{\eta}_k) \).

Case 2) Corresponds to the upper branch in (12) and lower branch in (13), and arises when \( \hat{z}_{k-1}/z_{ss} (\hat{\eta}_k) < 1 \) and \( \Delta z_k \) is sufficiently large so that \( \left( \hat{z}_{k-1} + \hat{\eta}_k \right)/z_{ss} (\hat{\eta}_k) \geq 1 \). This case can occur, for example, with high velocity or with a long interval between samples. Taking the lower branch in (13) is equivalent to \( \Delta r_k/L_1 = 1 \), where limit \( L_1 \) is given by

\[
L_1 = \frac{z_{ss} (\hat{\eta}_k) - \hat{z}_{k-1}}{z_{ss} (\hat{\eta}_k) - z_{ss} (\hat{\eta}_k) - \alpha \left( \hat{z}_{k-1}, \hat{\eta}_k \right) \hat{z}_{k-1}}.
\]

In this case \( \eta_k \) is computed in (12), but the saturation operates in (13) to give \( \hat{z}_k = z_{ss} (\hat{\eta}_k) \).

Case 3) Corresponds to the upper branches of (12) and (13), and arises when \( \hat{z}_{k-1}/z_{ss} (\hat{\eta}_k) < 1 \) and \( \Delta r_k/L_1 \leq 1 \). In this case the discrete-time elasto-plastic model approximates the continuous time model.

Through the operation of \( \alpha (z, \dot{z}) \) in (10) and state update (12) and (13), the elasto-plastic model has two characteristics that
are established here, and used to demonstrate the properties as follows.

\[ 0 \leq y_k \text{sgn}(\tau_k) \leq \Delta \tau_k \text{sgn}(\tau_k), \quad (15) \]

**Proof:** Proof is given for \( \tau_k > 0 \). For \( \tau_k < 0 \) follows with suitable changes of sign. On the lower branch of (12), \( y_k = 0 \) and so (15) is verified. On the upper branch and with \( \tau_k > 0 \), (15) becomes
\[
0 \leq \left[ 1 - \alpha \left( \hat{\tau}_{k-1}, \tau_k \right) \frac{\hat{\tau}_{k-1}}{\tau_k} \right] \leq 1.
\]

Recalling that \( \alpha(\cdot) \) is chosen so that \( 0 \leq \alpha(\hat{\tau}_{k-1}, \tau_k) \leq 1 \), the upper inequality is verified whenever \( \text{sgn}(\hat{\tau}_{k-1}) = \text{sgn}(\tau_k) \), but when \( \text{sgn}(\hat{\tau}_{k-1}) \neq \text{sgn}(\tau_k) \), \( \alpha(\hat{\tau}_{k-1}, \tau_k) = 0 \), and so the upper inequality is verified. The lower inequality is verified whenever \( \hat{\tau}_{k-1} \leq \tau_k \) (or \( \hat{\tau}_{k-1} > \tau_k \)), but when \( \hat{\tau}_{k-1} \geq \tau_k \), the lower branch of (12) is chosen, and so (15) is verified in all cases and the lemma is proven.

**Lemma 2:** (Plastic displacement occurs only in the motion direction) Given the EP friction model, (10), (12) and (13), a sequence of measurements \( \{ \tau_k \} \) and plastic displacement given by \( \Delta \hat{\tau}_k = \tau_k - \hat{\tau}_k \), then either \( \Delta \hat{\tau}_k = 0 \) or
\[
\text{sgn}(\Delta \hat{\tau}_k) = \text{sgn}(\Delta \hat{\tau}_k). \quad (16)
\]

**Proof:** Proof is given for \( \tau_k > 0 \). Proof for \( \tau_k < 0 \) follows with suitable changes of sign. In Case 1, \( \hat{\tau}_{k-1} > \tau_k \) (or \( \hat{\tau}_{k-1} \geq \tau_k \)), and so \( \Delta \hat{\tau}_k < 0 \) and \( \Delta \hat{\tau}_k = -\Delta \tau_k \). Recalling the requirement that \( \text{sgn}(\Delta \hat{\tau}_k) = \text{sgn}(\tau_k) \), it follows that (16) is verified in Case 1. In Case 2, \( 0 \leq \Delta \hat{\tau}_k \leq \tau_k \) where the last inequality follows from Lemma 1, and so \( 0 \leq \Delta \hat{\tau}_k \leq \Delta \hat{\tau}_k \), and either \( \Delta \hat{\tau}_k = 0 \) or (16) is verified. In Case 3, \( \Delta \hat{\tau}_k = \tau_k \), and so by Lemma 1 either \( \Delta \hat{\tau}_k = 0 \) or (16) is verified.

B. Demonstration of Four Properties of the EP Model

Using the cases and lemmas laid out previously, four properties of the EP model are established.

**Theorem 1:** (Boundedness of the state) Given the EP friction model, (8), (10), (12), and (13), a sequence of measurements \( \{ \tau_k \} \), and given \( |\hat{\tau}_k| \leq z_{\text{max}} \) for some \( j \geq 0 \), then \( |\hat{\tau}_k| \leq z_{\text{max}} \), for all \( k > j \).

**Proof:** The saturation operations in (12) and (13) assure that \( |\hat{\tau}_k| \leq z_{\text{max}} \), for all \( k \geq 1 \).

**Theorem 2:** (Convergence during steady sliding) Given the EP friction model, (8), (10), (12), and (13) and a sequence of measurements \( \{ \tau_k \} \) with \( \tau_k \neq 0 \), then
\[
\lim_{k \to \infty} \hat{\tau}_k = z_{\text{max}}(\tau) .
\]

**Proof:** If \( \hat{\tau}_{k-1} = z_{\text{max}}(\tau) \) then the lower branch in (13) is taken, and \( \hat{\tau}_k = \hat{\tau}_{k-1} = z_{\text{max}}(\tau) \); thus if the update rule of Case 1 or Case 2 arises on any sample \( j \), then \( \hat{\tau}_k = z_{\text{max}}(\tau) \), for all \( k > j \). Case 3 is the remaining possibility of the state update. To establish that a sequence of samples with \( \hat{\tau}_k \) updated according to Case 3 converges to \( z_{\text{max}}(\tau) \), consider the dynamics of \( \epsilon_k \), where \( \epsilon_k = z_{\text{max}}(\tau) - \hat{\tau}_k \). From (12) and (13), one finds
\[
\epsilon_k = \frac{1 - \alpha(\hat{\tau}_{k-1}, \tau_k)}{z_{\text{max}}(\tau_k)} \epsilon_{k-1} + \frac{\alpha(\hat{\tau}_{k-1}, \tau_k)}{z_{\text{max}}(\tau_k)} = 1, \quad \epsilon_{k-1} = 0
\]

In steady sliding, \( \Delta z_{\text{max}}(\tau) = 0 \). Because \( \epsilon_{k-1} = 0 \) is a fixed point of (17). The fixed point is stable if
\[
-1 \leq 1 - \alpha(\hat{\tau}_{k-1}, \tau_k) \frac{\Delta \hat{\tau}_k}{z_{\text{max}}(\tau_k)} \leq 1.
\]

The extremal value for the center term of (18) while remaining in Case 3 is given when \( \Delta \hat{\tau}_k = L_3 \). Substituting \( L_3 \) for \( \Delta \hat{\tau}_k \) in (18), with some manipulation gives the requirements that
\[
-1 + \alpha(\hat{\tau}_{k-1}, \tau_k) \frac{\Delta \hat{\tau}_k}{z_{\text{max}}(\tau_k)} \leq 1 - \alpha(\hat{\tau}_{k-1}, \tau_k)
\]

which is also verified by the properties of \( \alpha(\hat{\tau}_{k-1}, \tau_k) \) and \( \hat{\tau}_k/z_{\text{max}}(\tau_k) \). Equality in (21) corresponds to marginal stability of the fixed point. Strict inequality, corresponding to \( \alpha(\hat{\tau}_{k-1}, \tau_k) > 0 \) and \( \hat{\tau}_{k-1} > z_{\text{max}}(\tau_k) \), assures asymptotic stability.

To establish convergence of all sequences \( \{ \hat{\tau}_k \} \) to \( z_{\text{max}}(\tau_k) \), it remains to show that \( \hat{\tau}_k \) enters the region on which \( \alpha(\hat{\tau}_{k-1}, \tau_k) > 0 \) from any initial condition \( \hat{\tau}_0 \). Considering (10), when \( \hat{\tau}_{k-1} < z_{\text{max}} \), \( \alpha(\hat{\tau}_{k-1}, \tau_k) = 0 \) and \( \Delta \hat{\tau}_k = -\Delta \tau_k \). The system enters the condition \( \alpha(\hat{\tau}_{k-1}, \tau_k) > 0 \) in \( j \) steps, where \( j \) is the smallest integer such that \( \sum_{k=0}^{j} \Delta \tau_k > (\hat{\tau}_{k-1} - \hat{\tau}_0) \). Thus, in the case that the system is in the region where \( \alpha(\hat{\tau}_{k-1}, \tau_k) = 0 \) (marginal convergence), \( \Delta \tau_k = -\Delta \tau_k \) until \( \alpha(\hat{\tau}_{k-1}, \tau_k) > 0 \) (asymptotic convergence) is obtained. If \( \Delta \tau_k < L_3 \) for all samples, the system remains in Case 3 and \( \hat{\tau}_k \) asymptotically approaches zero. If there is a sample \( k \in I \) in which \( \Delta \tau_k > L_3 \), the system moves to Case 2 and \( \hat{\tau}_k = 0 \), for all \( k > I \).

**Theorem 3:** (The friction force opposes plastic displacement) Given the EP friction model, (8), (10), (12), and (13) with \( f_f(\hat{\tau}_k) \) given by (14), a sequence of measurements \( \{ \tau_k \} \), and given that friction model parameter \( \sigma_1 \) is chosen according to
\[
\sigma_1 < \sigma_0 \frac{f_C}{f_{\text{max}} - f_C} + \sigma_2 \frac{\tau_k}{z_{\text{max}} - z_C}
\]

then
\[
f_f(\hat{\tau}_k) \Delta \hat{\tau}_k > 0 \quad \forall \Delta \hat{\tau}_k \neq 0
\]

where \( \hat{\tau}_k = \tau_k - \hat{\tau}_k \), and where \( \Delta \hat{\tau}_k = \hat{\tau}_k - \hat{\tau}_{k-1} \) is the plastic displacement of the EP model in sample \( k \). Differences between
The discrete and continuous-time bounds for $\sigma_1$ are addressed in Section IV-C.

Proof: By lemma 2, $\Delta 0_k = 0$ or $\text{sgn}(\Delta 0_k) = \text{sgn}(\Delta \tilde{t}_k)$, by the definition of $\alpha(\cdot)$, $\Delta \tilde{t}_k \neq 0$ implies that $\text{sgn}(\Delta \tilde{t}_k) = \text{sgn}(\Delta \tilde{t}_k)$, and so in the product

$$f_f(\tilde{t}_k) \Delta \tilde{w}_k = \sigma_0 \Delta \tilde{z}_k \Delta \tilde{w}_k + \sigma_1 \Delta \tilde{z}_k \Delta \tilde{w}_k + \sigma_2 \Delta \tilde{t}_k \Delta \tilde{w}_k \tag{24}$$

the first and third terms are assured to be positive. Additionally, by Lemma 1, the second term is positive in Cases 2 and 3.

The remaining challenge arises because $\text{sgn}(\Delta \tilde{z}_k) \neq \text{sgn}(\Delta \tilde{t}_k)$ in Case 1. For this case, the $\sigma_1$ term in (14) does not oppose plastic sliding, and it is necessary to show that this term is dominated by the other two. Factoring out the $\Delta \tilde{t}_k$ in (24), it must be shown that

$$\left(\frac{\sigma_0 \Delta \tilde{z}_k + \sigma_1 \Delta \tilde{t}_k}{\Delta \tilde{t}_k + \sigma_2 \Delta \tilde{t}_k}\right) \text{sgn}(\Delta \tilde{t}_k) > 0. \tag{25}$$

Taking advantage of the fact that the value of $\Delta \tilde{z}_k$ is known to be $\sigma_0 \Delta \tilde{z}_k$ in Case 1, and that the negative-most value possible for $\Delta \tilde{z}_k$, $\text{sgn}(\Delta \tilde{t}_k) = \Delta \tilde{z}_k$, $\text{sgn}(\Delta \tilde{t}_k) = (f_c - f_{\text{max}})/\sigma_0$, the bound for $\sigma_1$ to assure (23) is given by (25).

Theorem 4: (The model is dissipative with respect to storage function $W(k)$, for all $\Delta \tilde{t}_k \neq 0$) Given the EP friction model, (8), (10), (12), and (13) with $f_f(\tilde{t}_k)$ given by (14), a sequence of measurements $\{\tilde{t}_k\}$, storage function $W(k) = \sigma_0 \Delta \tilde{z}_k^2/2$ and given that friction model parameter $\sigma_1$ is chosen according to (22), above, then for all $\Delta \tilde{t}_k \neq 0$

$$f_f(\tilde{t}_k) > W(k) - W(k-1). \tag{26}$$

Proof: Using $\Delta \tilde{t}_k + \Delta \tilde{w}_k$, (26) can be expanded to give

$$\sigma_0 \Delta \tilde{z}_k (\Delta \tilde{z}_k + \Delta \tilde{w}_k) + \Delta \tilde{t}_k \left(\frac{\sigma_1 \Delta \tilde{z}_k}{\Delta \tilde{t}_k + \sigma_2 \Delta \tilde{t}_k}\right) > \sigma_0 \Delta \tilde{z}_k \Delta \tilde{z}_k - \frac{1}{2} \sigma_0 \Delta \tilde{z}_k^2$$

which gives

$$\sigma_0 \Delta \tilde{z}_k \Delta \tilde{w}_k + \frac{\Delta \tilde{t}_k \sigma_1 \Delta \tilde{z}_k}{\Delta \tilde{t}_k} + \Delta \tilde{t}_k \sigma_2 \Delta \tilde{t}_k + \frac{1}{2} \sigma_0 \Delta \tilde{z}_k^2 > 0. \tag{27}$$

By lemma 2, $\Delta \tilde{w}_k = 0$ or $\text{sgn}(\Delta \tilde{w}_k) = \text{sgn}(\Delta \tilde{w}_k)$, and by the choice for $\alpha(\cdot)$, $\Delta \tilde{w}_k = 0$ when $\text{sgn}(\Delta \tilde{z}_k) = \text{sgn}(\Delta \tilde{w}_k)$, and so the first term makes a non-negative contribution. The third term is assured to be positive by the requirement that $\text{sgn}(\Delta \tilde{t}_k) = \text{sgn}(\Delta \tilde{t}_k)$ and the fourth term is non-negative. The second term is now to be considered.

In Cases 2 and 3, either $\Delta \tilde{z}_k = 0$ or $\text{sgn}(\Delta \tilde{z}_k) = \text{sgn}(\Delta \tilde{t}_k)$, and so the second term makes a non-negative contribution in these cases. It remains to show that the second term is dominated by the first and third terms in case 1 when $\sigma_1$ satisfies (22). Exploiting the fact that in case 1 $\text{sgn}(\Delta \tilde{z}_k) \neq \text{sgn}(\Delta \tilde{t}_k)$, and so $\sigma_0 \Delta \tilde{z}_k \Delta \tilde{w}_k > 0$, $\Delta \tilde{t}_k \cdot \text{sgn}(\Delta \tilde{z}_k)$, which gives

$$\sigma_0 \Delta \tilde{z}_k \Delta \tilde{w}_k + \frac{\Delta \tilde{t}_k \sigma_1 \Delta \tilde{z}_k}{\Delta \tilde{t}_k} + \Delta \tilde{t}_k \sigma_2 \Delta \tilde{t}_k + \frac{1}{2} \sigma_0 \Delta \tilde{z}_k^2 > 0.$$ 

The right inequality in (27) is demonstrated by (25). It is established when $\sigma_1$ satisfies (22), which completes the proof.

C. Discussion

The bound on $\sigma_1$ given in (22) is the discrete equivalent of the bound given in [11] and [18] for the continuous-time state-variable friction model, which can be written

$$\sigma_1 < \sigma_2 f_c / (f_{\text{max}} - f_c).$$

The bound on $\sigma_1$ for the discrete model, however, has a substantially different form from that found for the continuous-time model. The differences between the continuous- and discrete-time bounds arise from the facts that $f_f(\tilde{t}_k)$ is computed only at sample instants $\{t_k\}$, the presence of the term with $\sigma_1$ in (22) and the fact that the allowable $\sigma_1$ increases with increasing $\Delta \tilde{t}_k$.

With theorems 1–4, several formal properties of the EP model are established. However, perhaps the most important property of the EP model is that its performance degrades gracefully as step size is increased, sample rate is reduced or noise is present in samples $\{\tilde{t}_k\}$. In Section V numerical examples of the EP model are shown, specifically exploring these characteristics.

V. ESTIMATOR PERFORMANCE

As a result of the time-free formulation, spatial discretization, not time discretization has an effect on the model fidelity. A practical consequence is that sensing resolution drives the observation fidelity. Considering the physical scales of presliding and Stribeck effects for metal to metal contact [1], [3], position must be estimated or sensed with a resolution one micron or better for presliding displacement to be well represented. Similarly velocity must be estimated or sensed with a resolution of at least 0.001 m/s for Stribeck friction to be well captured. Again, this figure presents a considerable instrumentation challenge. At a rate of 1 kHz, if we set the velocity resolution quantum $\Delta v/\Delta T$ at a 1/10 of this value, the position should be resolved at 100 nm! These numbers indicate the use of down-sampling techniques for velocity estimation [19]. As a matter of fact, today’s industrial practice in machine tool applications calls for 50-nm resolution in the detection of slideway movements. As these values are the physical scales of the phenomena, these stringent sensing or estimation requirements apply to any friction estimator. To illustrate these issues, the estimator’s performance with respect to discretization resolution, initial conditions and noisy inputs is detailed below for a simple example. A system comprising a mass sliding on a frictional surface under the action of an applied force, is simulated using a stiff ODE solver to approach the conditions of continuous time. The trajectory $x(t)$ is recorded during simulation and supplied as an input to the discrete friction estimator. An update, according to (13), is triggered when an incremental displacement is detected to be of a magnitude larger than a given resolution $\Delta x$. The Stribeck curve is described by (8), the parameters are given in Table II.

A. Fidelity

Estimation results for several values of spatial resolution $\Delta x$ are compared with the continuous time solution using the applied force shown in Fig. 6(a) to illustrate the various regimes. The estimates of elastic displacement $\tilde{z}_k$ and friction
TABLE II
FRICTION MODEL PARAMETERS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>10.0 kg</td>
<td>$\sigma_0$</td>
<td>$1.0 \times 10^5$ N/m</td>
</tr>
<tr>
<td>$f_C$</td>
<td>1.0 N</td>
<td>$\sigma_1$</td>
<td>$1.0 \times 10^5$ N/s/m</td>
</tr>
<tr>
<td>$f_{\text{max}}$</td>
<td>1.5 N</td>
<td>$\sigma_2$</td>
<td>$4.0 \times 10^{-1}$ N/s/m</td>
</tr>
<tr>
<td>$v_S$</td>
<td>$10^{-3}$ m/s</td>
<td>$\delta_{\text{th}}$</td>
<td>$8.0 \times 10^{-6}$ m</td>
</tr>
</tbody>
</table>

Fig. 6. Numerical behavior of the estimator. The solid curve shows the applied force and reference response, the dot and dash curves show the response of the discrete-time EP model for three levels of spatial quantization.

force $\hat{f}_f(t_k)$ are shown in Fig. 6(b) and (c), respectively, for three values of spatial resolution, $Dx$. As the spatial resolution becomes smaller, the estimation approaches the continuous case.

An initial force spike causes transient sliding, then a large magnitude oscillating force causes the contact to transition from presliding to sliding several times. The applied force then returns to zero allowing the contact to settle unforced. At the highest resolutions 0.1 and 1 $\mu$m, the responses are not graphically distinguishable from the continuous case. Detailed views of three segments of the system response are seen in Fig. 7. In Fig. 7, at the highest resolution $Dx = 0.1 \mu$m, the estimate is still not graphically discernible from the continuous $z(t)$. The observation error degrades gracefully for an increase in $Dx$ covering nearly two orders of magnitude, up to 50% of $\inf \hat{z}_k(z_{\text{ref}}) = f_C/\sigma_0$. Recalling the convergence condition (18), at the lowest resolution, the discrete updates become visible. In Fig. 7(b), the region starting at $t_k = 0.6$ s shows the transition from presliding to sliding and back to presliding. Errors are well behaved for all step sizes, and, as expected, the response of the discrete-time implementation approaches the behavior of the continuous-time model as $Dx$ grows small relative to $z_{\text{ref}}$.

Fig. 7(c) shows the contact settling when the applied force becomes zero at $t_k = 1.0$ s. For relatively large $Dx$ the discrete-time model shows a steady-state error. This is expected because of the absence of time-varying signals to activate the friction model dynamics. The final error value is bounded by the spatial resolution.

**B. Convergence**

To illustrate state convergence, an incorrect initial condition is applied for the simulations of Figs. 6 and 7. For these figures, the discrete-time model is initialized to $\hat{z}_k = -10,0$, where as the reference model is initialized to $z(t) = 0$. From theorem 2, state observation error is expected to go to zero in steady sliding. The reduction in estimator error is seen in Fig. 6(b) and (c), and especially in the enlargement of Fig. 7(a). The figures show the estimation error to go to zero for all levels of spatial resolution. Note that some degree of plastic displacement is necessary to reduce estimation error, since during purely elastic displacement,
displacement and velocity measurements alone are insufficient to estimate \( z(t) \).

### C. Effect of Noise

During elastic presliding, the incremental uncertainty in \( z \) is at most the incremental uncertainty in \( x \), irrespective of the model parameters. Two cases arise. When the noise amplitude is smaller than sampling interval \( \Delta x \), the presence of noise is reflected as small errors in the state update, but when the noise amplitude is greater than \( \Delta x \), spurious sliding occurs, resulting in larger errors. This behavior is seen in Fig. 8, obtained by adding uniform random noise of magnitude 1 \( \mu m \) to the trajectory \( x(t) \) supplied as an input to the friction estimator. Comparison with Fig. 7 shows that for a \( \Delta x \) larger than or equal to the noise amplitude, the prediction is only marginally affected, but for a \( \Delta x \) smaller than the noise (\( \Delta x = 0.1 \mu m \) in the figure), the prediction error becomes larger than the noise. We can observe the beneficial effect of the correct selection the estimator spacial sampling relatively to the quality of the sensor signal.

### VI. Conclusion

The Dahl-type friction models offer a good balance between ease of implementation and fidelity to the details of friction, with guaranteed boundedness and convergence. With the addition of a suitably designed function \( \alpha(z, \dot{z}) \) to control plastic displacement, static friction is modeled and drift is avoided. The later characteristic is particularly useful for applications where the response of the model to sensor noise must be considered. The model also can represent a simplified form of friction hysteresis. The discrete-time implementation has been shown to be stable and converge to the desired steady-state value in steady sliding, independent of the sample size. And although the highest fidelity to the continuous time model is achieved with low noise and small steps, examples show the estimation to degrade gracefully as noise and step size are increased under the simple condition that the step size should always be greater or equal to the noise.

Further research could be pursued in several directions. There is some evidence that single-state models may be applicable to the cases where transmissions are considerably more elastic than contacts [20], whereas for machine tools of primary focus in this article, the opposite is generally true. Also, there might be advantages in using integration schemes that are more sophisticated than Euler integration. Can the performance guarantees of theorems 1–4 be established for a more sophisticated integrator? Finally, are there methods to convert continuous multistate friction models into their discrete counterparts provably having equivalent properties?

### REFERENCES