A reduced-order model from high-dimensional frictional hysteresis

Saurabh Biswas and Anindya Chatterjee

Department of Mechanical Engineering, IIT Kanpur, Kanpur,
Uttar Pradesh 208016, India

Hysteresis in material behaviour includes both signum nonlinearities as well as high dimensionality. Available models for component-level hysteretic behaviour are empirical. Here, we derive a low-order model for rate-independent hysteresis from a high-dimensional massless frictional system. The original system, being given in terms of signs of velocities, is first solved incrementally using a linear complementarity problem formulation. From this numerical solution, to develop a reduced-order model, basis vectors are chosen using the singular value decomposition. The slip direction in generalized coordinates is identified as the minimizer of a dissipation-related function. That function includes terms for frictional dissipation through signum nonlinearities at many friction sites. Luckily, it allows a convenient analytical approximation. Upon solution of the approximated minimization problem, the slip direction is found. A final evolution equation for a few states is then obtained that gives a good match with the full solution. The model obtained here may lead to new insights into hysteresis as well as better empirical modelling thereof.

1. Introduction

Hysteresis, a phenomenon involving persistent memory effects, occurs in many systems. The observation of hysteresis in magnetism is attributed to Ewing [1], and a nice stress–strain hysteresis curve for material dissipation was given by Rowett [2]. An early mathematical model for hysteretic behaviour in ferromagnetism was given by Preisach [3]. More recently, hysteresis has been noted in many other areas, and there has been much research in the physics, modelling,
Figure 1. Hysteresis loops in mechanical elements (a) and magnets (b). $F$ is a force and $x$ is a displacement or stretching response; $H$ is an applied magnetic field and $B$ is the resulting magnetic flux density (the response). In (a), the steady periodic response traverses the loop clockwise and in (b), it is anticlockwise. There can be transients before the steady-state behaviour is established, though they have not been indicated in the figures, for simplicity.

numerical simulation and dynamics of systems with hysteresis (see [4] and references therein). To depict key qualitative features, a hysteresis curve for a mechanical element is sketched in figure 1a, and another for a magnet is sketched in figure 1b.

The underlying microscopic physics of material hysteresis is complicated. Attempts have been made over several decades to describe hysteretic responses by means of physical, semi-physical and purely empirical models [4]. A dominant majority of theoretical papers on hysteretic phenomena have concerned themselves with magnetism, for which complex theories, models and simulations have been developed.

However, options have been limited for simple simulation of hysteretic systems using modest computational effort. Prominent among these have been the classical and modified Preisach models and, in structural mechanics, the Bouc–Wen model [5,6]. Of these, the Preisach-type models are somewhat complicated for numerical use. These models work with an underlying distribution of idealized hysteretic elements, and during loading and unloading the states of these distributed elements need to be tracked. The tracking is done using vertical and horizontal lines that sweep out portions within a triangular region where the random parameters of the hysteresis elements are distributed. The Preisach model’s evolution is thus given in terms of successive geometrical constructions and is not easily expressible in terms of differential equations. In contrast to the Preisach-type models, the Bouc–Wen model involves a single scalar differential equation and is much easier to use, but has a fundamental limitation as we will describe below.

Our own study of hysteresis models is motivated by an interest in internal damping in materials [7,8]. For modelling internal damping under temporally complex stress cycles (specifically with minor unloading loops within larger loading cycles), we seek appropriate low-dimensional empirical models of hysteretic response and dissipation.

For hysteretic behaviour in structural mechanics, the Bouc–Wen model has been used by many authors (see [9] and references therein). In the basic Bouc–Wen model, the hysteretic part of the force is given by a single internal variable $z$ driven by a displacement input $u$, as in

$$
\dot{z} = \tilde{A} \dot{u} - \tilde{a} |\dot{u}| z |z|^{n-1} z - \tilde{\beta} \dot{u} |z|^n, \quad (1.1)
$$

1User-friendly code for the Preisach model is available in Bertotti & Mayergoyz [4], vol. I, p. 683. There, evolution equations are not solved explicitly: the forcing history of interest is to be supplied in advance.
Figure 2. Response of the Bouc–Wen model. Here, $\bar{A} = 1$, $\bar{\alpha} = 0.8$, $\bar{\beta} = 0.5$ and $n = 2$. Here, $u$ was chosen to ensure that $z = 0.619 (\sin(t) + 0.4 \sin(4.16753t))$; these numerical values will be re-used later, but are otherwise arbitrary. Inset: a sketch of the sort of minor loop closure that is not shown by the Bouc–Wen model. (Online version in colour.)

where $\bar{A}$, $\bar{\alpha}$, $\bar{\beta}$ and $n$ are parameters that must satisfy

$$\bar{A} > 0, \; \bar{\alpha} > 0, \; \bar{\beta} \in [-\bar{\alpha}, \bar{\alpha}] \; \text{and} \; n > 0$$

[10,11]. A key aspect of hysteretic material behaviour is that, within a larger load cycle, if there is a minor unloading–loading loop, then the response also shows a minor loop that turns around and intersects itself. The Bouc–Wen model cannot incorporate such self-intersection or minor loop closure. Figure 2 illustrates such non-closure, which leads to large differences between simulated and physically relevant behaviours. As seen in the figure, the hysteresis loops slide around too much in the horizontal direction.

The inability of the Bouc–Wen model to capture such minor loops has motivated further ad hoc treatment (see reference [12]). However, we adopt a more constructive approach in this paper. To this end, note that the Bouc–Wen model (equation (1.1)) is a special case of a more general class of models of the form

$$\dot{z} = \dot{u} f(z, u, \text{sgn}(\dot{u})), \quad (1.2)$$

where the pre-multiplier $\dot{u}$ and the internal dependence only on the sign of $\dot{u}$ together ensure rate-independence (other examples may be found in references [13,14]). The above can be rewritten as

$$\frac{dz}{du} = f(z, u, \text{sgn}(\dot{u})). \quad (1.3)$$

If the right-hand side of equation (1.3) did not include dependence on $\text{sgn}(\dot{u})$, then loading and unloading paths would be the same, and there would be no hysteresis. In this way, the Bouc–Wen model is clever and simple. However, models of the form of equation (1.2) cannot capture minor loops if there are more than two rate reversals within one forcing cycle, as may be seen from figure 3. In figure 3, responses are shown to two different forcing histories: one of small amplitude, with no additional rate reversals within the forcing cycle, and one of larger amplitude, with additional rate reversals within the forcing cycle. The amplitudes are adjusted, so that the two solution loops touch at a point: it is emphasized that these are two different responses, under different forcing histories, of the same hypothetical hysteretic material. Such circumstances are
Two different solutions indicates internal variables (a)(b)

**Figure 3.** (a) Responses to two different hypothetical forcing histories. One has a small amplitude and two rate reversals per period (dashed line). The other has a larger amplitude with unloading or subloops within each cycle (solid line). The relative amplitudes of forcing can be adjusted to make the two solution curves touch at a point. (b) At the point where the two solution curves touch, there are two different tangents as indicated: here, $z$, $u$ and $\text{sgn}(\dot{u})$ are the same, but $dz/du$ differs, and so equation (1.3) is insufficient. (Online version in colour.)

It is easy to create in experiments with magnets [15], and our high-dimensional frictional model below will also show such solutions. There is nothing unphysical about the circumstances depicted in the figure.

Yet, at the point where the two solution loops touch, there are two different curves along which $F$ and $x$ both increase. Recall that $F$ and $x$ here correspond to $z$ and $u$ in equations (1.2) and (1.3). Thus, the right-hand side of equation (1.3) is the same, but the left-hand side is not. We conclude that physical behaviour as depicted in figure 3 implies hidden internal variables in addition to $z$; and this is a basic weakness of all models of the form of equation (1.2).

We can now motivate this paper as follows. Experiments with magnets or hysteretic mechanical elements, even under complex loading, often only give access to the external variables ($F$ and $x$; or $H$ and $B$; or more abstractly $z$ and $u$). By contrast, a numerical study of a frictional system with many internal variables can give us access to all internal variables. A study of such a system may yield better understanding of low-order modelling of hysteresis in general. In this paper, we present a study of such a high-dimensional frictional rate-independent hysteretic system, and show how a low-dimensional model can be purposefully constructed from it. We have not found any similar study in the literature.

## 2. High-dimensional frictional system

**Figure 4** schematically depicts a high-dimensional frictional system. Here $B_1$, $B_2$, $B_3$, ..., $B_N$ are $N$ massless blocks. These blocks are arbitrarily interconnected by springs of stiffness $k_1$, $k_2$, $k_3$, .... In **Figure 4**, $b_1f(t)$, $b_2f(t)$, $b_3f(t)$, ..., $b_Nf(t)$ are external forces on the blocks. Friction forces on the blocks are written as $F_1 = -\mu_1 \text{sgn}(\dot{x}_1)$, $F_2 = -\mu_2 \text{sgn}(\dot{x}_2)$, $F_3 = -\mu_3 \text{sgn}(\dot{x}_3)$, ..., $F_N = -\mu_N \text{sgn}(\dot{x}_N)$.

The governing equation is (see appendix A)

$$\mu \text{sgn}(\dot{x}) + Kx = bf(t), \quad (2.1)$$

where $x$ is an $N$-dimensional vector, $\mu$ is an $N \times N$ diagonal matrix with positive elements, $K$ is a symmetric positive definite matrix of size $N \times N$, $b$ is an $N$-dimensional vector and $f(t)$ is a scalar differentiable function of time representing an oscillating load; and where the signum function ‘$\text{sgn}$’ is defined elementwise and understood to be plus 1 for positive values of the argument, minus 1 for negative values of the argument, and multi-valued (within $[-1, 1]$) when the argument is zero.
Equation (2.1) is not a set of ordinary differential equations (ODEs) in the usual sense, because the signum function is not invertible. However, the system can be incrementally solved by casting it into the form of a linear complementarity problem (LCP; see [16]), as outlined in appendix A. For numerical solution of the LCP for each time increment, we used Lemke’s algorithm as implemented in a freely available Matlab program (see [17])2. For verification, we also solved the above system for moderate \( N \) using a smoothed version of the signum function along with some numerical tricks, and obtained the same results as from the quicker LCP; details are omitted.

(a) Direct numerical solution

We first solve equation (2.1) for random choices of \( \mu, K \) and \( b \), and under oscillatory \( f(t) \). To this end, we generate a 500-dimensional random system in Matlab as follows.3 The \( \mu \)-values are uniformly distributed in \((0, 1)\). \( K \) has random orthogonal eigenvectors and eigenvalues uniformly distributed in \((0, 3)\). The elements of \( b \) are normally distributed with zero mean and unit variance. Finally, we take \( f = \sin(t) + 0.4 \sin(4.16753t) \) for initial simulation (compare this \( f \) with \( z \) in figure 2). Figure 5 shows the obtained results. These are combined into an effective scalar displacement below.

(b) Underlying two-dimensional plot

Our \( f(t) \) is generalized scalar force, whereas the response \( x \) above is high dimensional. We need an effective scalar displacement, which we will identify using the work done by the external forces. The incremental work done on the system by all the external forces is

\[
dW = f b^T dx
\]

Letting \( b^T dx = d\xi \), or \( \xi = b^T x \), we find \( dW = f d\xi \). Thus, \( \xi \) is the appropriate generalized displacement. We can now study the hysteretic behaviour of the frictional system through plots of force \( f \) against displacement \( \xi \).

Figure 6 shows solutions for zero initial conditions and three different forcing histories: \( f, 0.65f \) and \( 0.8f \).

The hysteresis curves obtained in figure 6 exhibit both major and minor loops, unlike the Bouc–Wen model. The solution curves for amplitudes 1 and 3 touch at a point, as anticipated in figure 3, showing that such behaviour is physically permissible and that hysteresis models need additional internal variables. They cannot have the simple form of equations (1.2) and (1.3).

We now have a high-dimensional rate-independent system that shows hysteresis. The system violates no physical laws. Unlike typical experimental systems, here we have access to every

---

2’CompEcon Toolbox.’ At the time of writing, the code is also available at http://people.sc.fsu.edu/~jburkardt/m_src/lemke/lemke.m.

3The Matlab commands are ‘\( n = 500; \) \( P=\text{orth}(	ext{randn}(n)); \) \( D=\text{diag}(3*\text{rand}(n,1)); \) \( K=P^*D*P; \) \( b=\text{randn}(n,1); \) \( \mu=\text{diag}(\text{rand}(n,1)); \)’
Figure 5. Response versus time for a random 500-dimensional massless frictional system. The different graphs show the various elements of $x$ plotted against time. Periods of constancy (sticking) and variation (slipping) are seen for each element. (Online version in colour.)

Figure 6. Hysteresis loops for the 500-dimensional frictional system with different forcing amplitudes. Amplitude 1 (‘amp 1’) has $f = \sin(t) + 0.4 \sin(4.16753t)$; amplitudes 2 and 3 refer to $f = 0.65(\sin(t) + 0.4 \sin(4.16753t))$ and $f = 0.80(\sin(t) + 0.4 \sin(4.16753t))$, respectively. (Online version in colour.)

internal state. We are now ready to attack the main problem addressed in this paper, namely the development of a reduced-order model for this specific hysteretic system.

3. Reduced-order model

As noted above, equation (2.1) is not a system of ODEs. Galerkin projections do not seem feasible. To obtain lower-order equations, we use ideas of work and dissipation.
Let \( \dot{x} = v \) in equation (2.1). We seek \( v \) for given values of \( x, f(t) \) and \( \dot{f}(t) \).

The rate of frictional dissipation cannot exceed the rate of work done by external forces minus the rate of increase of system potential energy. Therefore, \( v^T b f(t) - v^T K x \) is the available dissipation budget, and \( v^T \mu \text{sgn}(\dot{x}) \) is the dissipation rate. For slip to occur,

\[
v^T \mu \text{sgn}(v) + v^T K x - v^T b f(t) \leq 0. \tag{3.1}
\]

If the inequality is strict, then there is an instant of infinitely fast slip, because the system has no inertia. For sustained slip at finite speeds under changing \( f \), equality must hold in equation (3.1).

In numerical work, we can minimize the left-hand side of equation (3.1) with respect to unit-norm \( v \) and see if the minimum is negative; if it is, slip occurs in the direction of \( v \).

We note that the quantity in equation (3.1) could be interpreted as a dissipation potential, with minor complications in that slip sometimes occurs and sometimes not. However, an exploration of the underlying theoretical and thermodynamic implications is not our goal here. We focus on model order reduction and refer the interested reader to, for example, Hackl & Fischer [18] and references therein.

To develop a reduced-order model, we need a set of basis vectors. For now, let

\[
x = Q q \tag{3.2}
\]

with \( Q^T Q = I \), where the columns of \( Q \) are yet to be chosen. We have

\[
v = Q q
\]

and hence

\[
v^T v = \eta^T Q^T Q \eta = \eta^T \eta.
\]

Substituting \( x = Q q \) and \( v = Q \eta \) into the left-hand side of equation (3.1) gives

\[
\eta^T Q^T \mu \text{sgn}(q) + \eta^T (Q^T K Q) q - \eta^T (Q^T b) f(t). \tag{3.3}
\]

We define

\[
\eta^T Q^T \mu \text{sgn}(q) = G(\eta), \quad Q^T K Q = \tilde{K} \quad \text{and} \quad Q^T b = \tilde{b}, \tag{3.4}
\]

so that equation (3.3) becomes

\[
G(\eta) + \eta^T (\tilde{K} q - \tilde{b} f(t)) = G(\eta) + \eta^T c, \tag{3.5}
\]

where in turn

\[
c = \tilde{K} q - \tilde{b} f(t). \tag{3.6}
\]

Our numerical approach involves, at each timestep, the minimization of

\[
G(\eta) + \eta^T c \tag{3.7}
\]

with respect to \( \eta \), for a known vector \( c \). However, because the quantity in equation (3.7) is homogeneous of degree one in \( \eta \), we use it only to identify the direction of slip, minimizing subject to \( \eta^T \eta = 1 \); the slip rate is specified later. A key point here is that the dimension of \( \eta \) equals the number of columns of \( Q \), and is much smaller than the \( N = 500 \) of the original model. Note that \( G(\eta) \) is complicated because of the signum of a high-dimensional vector that appears within it (equation (3.4)).

(a) Choice of basis vectors

We choose our basis vectors via the singular value decomposition (also known as the proper orthogonal decomposition; [19]) applied to the full response \( x \) computed as described above. In the foregoing three simulations, for example, the full model was 500-dimensional, and the number of timesteps taken was 1000. The 500-dimensional state (\( x \)-vector) obtained at each instant makes up one row of the data matrix. Hence, we obtain a \( 1000 \times 500 \) response matrix, which with a slight abuse of notation we also call \( x \) within the Matlab environment. The Matlab built-in...
Figure 7. First 20 singular values of the response $x$. (Online version in colour.)

function ‘svd’ gives proper orthogonal modes and singular values$^4$ of $x$. Figure 7 shows the first 20 singular values of the response matrix $x$. The rapidly decaying singular values suggest that a low-dimensional description of the data is feasible; we used the first three singular vectors (or proper orthogonal modes) for our reduced-order modelling. However, we found that small intervals of motion near turning points in the forcing require special attention for a good approximation. Accordingly, we also included three singular vectors corresponding to motions immediately following reversals in the forcing. We thus have six basis vectors which we orthogonalize to obtain $Q$.

(b) Approximation of $G(\eta)$

A key step now is approximation of $G(\eta)$ of equation (3.7) by an analytically tractable, smooth function. We note that $G(\eta)$ is homogeneous of degree one in $\eta$, and consider (somewhat arbitrarily)

$$G(\eta) = \frac{(\eta^T A \eta)^\beta}{(\eta^T \eta)^{\beta - 1/2}}.$$  \hspace{1cm} (3.8)

Here, $A$ is a symmetric positive definite matrix, to be fitted along with the scalar parameter $\beta$. The fitting of $A$, for given $\beta$, can be simplified if we transform equation (3.8) to

$$(G(\eta) \cdot (\eta^T \eta)^{\beta - 1/2})^{1/\beta} = \eta^T A \eta.$$  \hspace{1cm} (3.9)

We proceed as follows. First, $\beta$ is chosen. Then, a large number (we used one million) of random vectors $\eta$ are generated, not of unit norm (we took the elements to be normally distributed with zero mean and unit variance). Because $\mu$ and $Q$ are known, the left-hand side can be evaluated for each $\eta$. The right-hand side is linear in the elements of $A$. $A$ is $6 \times 6$ in our case, but is symmetric, so it has 21 independent elements. Thus, we get one million simultaneous equations in 21 variables (an overdetermined system). Solving these in a least-squares sense gives an estimate for $A$ for the chosen $\beta$.

$^4$The command ‘[u,s,v]=svd(x,0);’ returns three matrices, where (i) ‘s’ is a diagonal matrix of size $500 \times 500$. Its elements are the singular values, arranged in decreasing order. (ii) ‘u’ and ‘v’ are orthogonal matrices of sizes $1000 \times 500$ and $500 \times 500$, respectively (not including the optional argument of ‘0’ in the call to ‘svd’ would return a $1000 \times 1000$ matrix ‘u’ and a $1000 \times 500$ matrix ‘s’). The columns of ‘v’ are the singular vectors or proper orthogonal vectors. Successive columns of ‘v’ indicate directions with successively lower mean square variation in $x$, with the first few columns capturing most of the variation.
We estimate $A$ in this way for several values of $\beta$; and in each case, we compute the sum of squares of the error computed from equation (3.8), i.e. the sum of squares of
\[
G(\eta) - \frac{(\eta^T A \eta)^\beta}{(\eta^T \eta)^{\beta-1/2}}.
\]
That sum of squares is divided by the sum of squares of $G(\eta)$, in order to obtain an overall measure of the error in the fit. The result, plotted against $\beta$, is shown in figure 8. We find that the error is reasonably small, and $\beta = 0.5$ is a good choice because it is analytically most convenient. Thus, we replace equation (3.8) with
\[
G(\eta) \approx (\eta^T A \eta)^{1/2}, \quad (3.10)
\]
where we have emphasized that we have an approximate fit and not an exact match. This approximation does not assume $\|\eta\| = 1$.

Because equation (3.10) introduces an ad hoc approximation, we can no longer expect our reduced-order model to get better and better as we include more basis vectors in $Q$. However, the approximation seems unavoidable to us.

(c) Slip direction

The slip direction $\eta$ minimizes $G(\eta) + \eta^T c$ subject to $\eta^T \eta = 1$. The minimization involves a few small tricks, but the minimizer can be found unequivocally as shown below.

Introducing a Lagrange multiplier for the constraint, we first set to zero the gradient of $G(\eta) + \eta^T c - \lambda \eta^T \eta$ with respect to $\eta$:
\[
\frac{\partial}{\partial \eta} \left\{ \sqrt{\eta^T A \eta + \eta^T c - \lambda \eta^T \eta} \right\} = 0,
\]
or
\[
\frac{1}{\sqrt{\eta^T A \eta}} A \eta + c - 2\lambda \eta = 0.
\]
Letting $2\lambda = \bar{\lambda}$,
\[
\frac{1}{\sqrt{\eta^T A \eta}} A \eta - \bar{\lambda} \eta + c = 0. \quad (3.11)
\]
$A$ is now factored using $A \Phi = \Phi \Lambda$, where $\Phi$ is an orthogonal matrix of eigenvectors and $\Lambda$ is a diagonal matrix of eigenvalues.
Let
\[ \eta = \Phi w. \quad (3.12) \]

Equation (3.11) gives
\[ \frac{A}{\sqrt{w^T A w}} w - \lambda w + \Phi^T c = 0. \quad (3.13) \]

Let \( \Phi^T c = \tilde{c} \) and \( 1/\sqrt{w^T A w} = \gamma \). Then,
\[ (\gamma A - \tilde{\lambda} I)w = -\tilde{c}. \]

Because \( \tilde{\lambda} \) has no special significance, we let \( \tilde{\lambda} = \gamma \hat{\lambda} \) to obtain
\[ \gamma (A - \hat{\lambda} I)w = -\tilde{c}. \]

Hence,
\[ w_i = \frac{1}{\gamma} \frac{\tilde{c}_i}{(\hat{\lambda} - \Lambda_i)} \quad (3.14) \]

Now, \( \eta^T \eta = 1 \) gives \( w^T \Phi^T \Phi w = w^T w = 1 \). Accordingly,
\[ w_i^2 = \frac{1}{\gamma^2} \frac{\tilde{c}_i^2}{(\hat{\lambda} - \Lambda_i)^2}. \quad (3.15) \]

Multiplying both sides by \( \Lambda_i \) and adding gives
\[ \sum_i w_i^2 \Lambda_i = \frac{1}{\gamma^2} \sum_i \tilde{c}_i^2 \Lambda_i. \]

By definition of \( \gamma \),
\[ \frac{1}{\gamma^2} = \sum_i w_i^2 \Lambda_i, \]
so
\[ \sum_i \frac{\tilde{c}_i^2 \Lambda_i}{(\hat{\lambda} - \Lambda_i)^2} = 1. \quad (3.16) \]

Equation (3.16) must be solved for \( \hat{\lambda} \). Assuming that the \( \tilde{c}_i \) are not all zero, it must have at least two real roots. To see this, note that, as \( \hat{\lambda} \) approaches the eigenvalues of \( A \) (\( \Lambda_i \), all real and positive), the left-hand side must go to infinity (i.e. have a pole) wherever the corresponding \( \tilde{c}_i \neq 0 \). However, to the right of the rightmost pole, the left-hand side of equation (3.16) goes to zero as \( \hat{\lambda} \to \infty \); and so it must equal 1 for some real \( \hat{\lambda} \) to the right of the rightmost pole. By similar arguments, there must be a real solution to the left of the leftmost pole. There may be other real solutions, occurring in pairs, between poles.

To solve equation (3.16), we combine the terms on the left-hand side by multiplying the individual denominators into a common denominator; and then we cross multiply with that common denominator to get a polynomial equation in \( \hat{\lambda} \) of order \( 2n \), where \( n \) is the size of \( A \) (as mentioned above, \( n = 6 \) for our reduced-order model). The polynomial to be solved is
\[ \sum_i \left( \tilde{c}_i^2 \Lambda_i \cdot \prod_{j \neq i} (\hat{\lambda} - \Lambda_j)^2 \right) = \prod_{i} (\hat{\lambda} - \Lambda_i)^2. \]

Computation of the coefficients of the polynomial, as well as finding its roots, can both be done numerically within Matlab; manual intervention is not required.
Solving equation (3.16) as above to find \( \hat{\lambda} \), we retain only the real roots. For each such root, using \( ||w|| = 1 \), equation (3.15) gives

\[
\sum_{i} w_i^2 = \frac{1}{\gamma^2} \sum_{i} \frac{\bar{c}_i^2}{(\hat{\lambda} - \Lambda_i)^2} = 1,
\]

whence we find

\[
\gamma = \sqrt{\sum_{i} \frac{\bar{c}_i^2}{(\hat{\lambda} - \Lambda_i)^2}} \tag{3.17}
\]

Equation (3.17) gives \( \gamma \) for each real \( \hat{\lambda} \). We can then find \( w \) using equation (3.14). Finally, \( \eta \) is obtained from equation (3.12). Among the finite number of candidate unit-norm \( \eta \)'s thus obtained, we choose by direct evaluation the one which minimizes \( \sqrt{\eta^T A \eta + \eta^T c} \).

In this way, given \( A \) and \( c \) (see equation (3.6)), we find the preferred slip direction \( \eta \).

(d) Reduced-order model

The unit vector \( \eta \) found above gives the direction of slip, should slip occur. The rate of slip remains to be found. In this sense, we are guilty of a small abuse of notation in that \( \eta \) is no longer \( \dot{q} \), but actually just gives the direction of \( \dot{q} \).

There are two approximations involved so far. One is in the small number of basis vectors chosen in \( Q \) (here, 6). The other is in the approximation of \( G(\eta) \) in equation (3.10). The first approximation can be systematically improved by increasing the number of basis vectors retained. The second approximation is fortuitous, \textit{ad hoc}, and not amenable to systematic improvement. We now wish to see the quality of approximations obtained.

Let \( y = \sqrt{\eta^T A \eta + \eta^T c} \), where \( c = \bar{K}q - \bar{b}f(t) \) and \( \eta \) is as found above. Recall that if \( y < 0 \), then slip should be infinitely fast; if \( y > 0 \), then slip should not occur; and if \( y = 0 \), then sustained finite-rate slip is possible. Given the approximations already made, within our goals of demonstrating a lower-dimensional model, and in the interest of simplicity, we adopt the following approximated evolution equation for the state \( q \) (recall equation (3.2)):

\[
\frac{dq}{dt} = \text{gain} \cdot \eta \cdot |y|^\alpha \cdot |\dot{f}| \cdot \{y < 0\}, \tag{3.18}
\]

where ‘gain’ is an arbitrary large positive multiplier (we used numbers in the range 1000–10 000); \( \alpha \) is a positive number (we used 1 or 1.2); the multiplier \( |\dot{f}| \) makes the reduced system rate-independent; and the multiplier \( \{y < 0\} \) is a logical variable (1 if the inequality holds, and 0 otherwise). The above evolution equation causes rapid slip, if \( y \) is large and negative; little slip, if \( y \) is very small in magnitude; and (because of the large gain) finite-rate slip, if \( y \) is negative and fairly small. Letting the gain approach infinity would, in the limit, give us finite rate slip as \( y \to 0^- \).

Equation (3.18) thus provides a reasonable approximation to the originally envisaged dynamics. In future work, a better numerical strategy may perhaps be developed, but this suffices for now.

For completeness, we report here the numerical values obtained for matrices \( A, \bar{K} \) and \( \bar{b} \) from the above-mentioned calculations. The original \( \mu, K \) and \( b \) are now gone; these three matrices
### Table 1. Load cases with varying amplitude.

<table>
<thead>
<tr>
<th>case</th>
<th>forcing (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4 sin(t) + sin(4.16753t)</td>
</tr>
<tr>
<td>2</td>
<td>sin(t) + 0.4 sin(4.16753t)</td>
</tr>
<tr>
<td>3</td>
<td>sin(t) + 1.5 sin(4.16753t)</td>
</tr>
<tr>
<td>4</td>
<td>1.5 sin(t) + sin(4.16753t)</td>
</tr>
<tr>
<td>5</td>
<td>sin(t) + 3.0 sin(4.16753t)</td>
</tr>
<tr>
<td>6</td>
<td>3.0 sin(t) + sin(4.16753t)</td>
</tr>
</tbody>
</table>

### Table 2. Load cases with varying frequency.

<table>
<thead>
<tr>
<th>case</th>
<th>forcing (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sin(t) + 1.5 sin(π/4 t)</td>
</tr>
<tr>
<td>2</td>
<td>sin(t) + 1.5 sin(π/2 t)</td>
</tr>
<tr>
<td>3</td>
<td>sin(t) + 1.5 sin(π t)</td>
</tr>
<tr>
<td>4</td>
<td>sin(t) + 1.5 sin(5π/4 t)</td>
</tr>
<tr>
<td>5</td>
<td>sin(t) + 1.5 sin(3π/2 t)</td>
</tr>
<tr>
<td>6</td>
<td>sin(t) + 1.5 sin(2π t)</td>
</tr>
</tbody>
</table>

The additional parameters ‘gain’ and $\alpha$ of equation (3.18) are part of the numerical solution strategy, and not a part of the reduced model. Below define the reduced-order model. The additional parameters ‘gain’ and $\alpha$ of equation (3.18) are part of the numerical solution strategy, and not a part of the reduced model.

$$A = \begin{bmatrix} 69.7660 & -16.7609 & 5.0974 & -0.4140 & 6.0795 & 0.0735 \\ 54.7643 & 21.1285 & -11.5050 & 9.9094 & 1.1897 \\ 53.1705 & 29.9240 & -1.0899 & 1.6273 & \end{bmatrix}, \quad (3.19)$$

$$\bar{K} = \begin{bmatrix} 0.3887 & 0.5313 & -0.1284 & 0.0080 & -0.0117 & 0.0328 \\ 1.2884 & -0.5771 & -0.0405 & -0.3183 & 0.0362 & \\ 1.3397 & -0.9649 & 0.1496 & 0.0970 & \end{bmatrix}, \quad (3.20)$$

$$\bar{b} = \begin{bmatrix} -25.6657 \\ -25.5468 \\ 8.7604 \\ -0.9557 \\ 1.6718 \\ 0.1085 \end{bmatrix}, \quad (3.21)$$

All reduced-order model simulations presented below are obtained using the matrices given by equations (3.19)–(3.21). These matrices represent the fitted parameters in a six-state model that approximates the original 500-state system. Admittedly, the number of fitted parameters is large ($21 + 21 + 6 = 48$ for $A$, $\bar{K}$, and $\bar{b}$, respectively). By a further coordinate transformation that diagonalizes $\bar{K}$, we can transform the reduced model to an equivalent one with fewer parameters ($21 + 6 + 6 = 33$), but we avoid that additional step, for brevity, in the presentation. Our main goal here has been systematic model order reduction, which has been achieved; and our hope is that...
Figure 9. Hysteresis curves for the forcing histories in table 1. Enlarged views are given in the electronic supplementary material. (Online version in colour.)

...this reduced model, in turn, provides insights that lead to more compact reduced-order hysteresis models with fewer parameters in future.

(e) Summary

Our reduced-order model is now summarized. Given $A$, $\bar{K}$, $\bar{b}$ matrices (equations (3.19)–(3.21)) and forcing $f(t)$ and current state $q$, we find an $\eta$ that minimizes

$$\min_{||\eta||=1} \sqrt{\eta^T A \eta + \eta^T (\bar{K}q - \bar{b}f)}.$$

Matlab code for finding the $\eta$ is given in the electronic supplementary material. For that $\eta$, we find

$$y = \sqrt{\eta^T A \eta + \eta^T (\bar{K}q - \bar{b}f)}.$$

Then, we solve

$$\frac{dq}{dt} = \text{gain} \cdot \eta \cdot |y|^{\alpha} \cdot |\dot{f}| \cdot \{y < 0\},$$
where ‘gain’ is a large multiplier (we used values between 1000–10 000) and $\alpha$ is a somewhat arbitrary positive parameter (we used values between 1 and 1.2).

We now present numerical results obtained using this reduced-order model.

4. Results and discussion

All our simulations of the full model given below were conducted using the LCP formulation described earlier; and all simulations of the reduced model were performed using equation (3.18). Because equation (3.18) is a stiff system, we used ‘ode15s’ (a stiff system solver) in Matlab. Tables 1 and 2 show the forcing histories ($f(t)$) used. Because the system response is rate-independent, the actual waveforms within $f$ are irrelevant; only the turning points or rate reversals matter. In the forcing histories considered, we include a fixed pair of frequencies with different relative amplitudes (table 1) as well as different pairs of frequencies with a fixed set of amplitudes (table 2).

Figure 9 shows the solutions obtained for the forcing histories in table 1. In these plots, $f$ was a known input for both full and reduced simulations. The generalized displacement $\xi$ was
taken to be $b^T x$ for the full simulation, and $\bar{b}^T q$ for the reduced simulation. It is seen that for a variety of combinations of forcing amplitudes, with small and large unloading loops, the reduced model does an excellent job of approximating the full system. A small issue in simulations of the reduced-order model is that we cannot start\(^5\) with both $q = 0$ and $f = 0$, and so the simulations were started with extremely small non-zero initial conditions for $q$.

Similarly, figure 10 shows results for the forcing histories in table 2. It is seen that changing the frequency combinations has no significant effect: the match remains good overall.

A minor flaw remains in the reduced model solutions. At points of load reversal, there is a brief interval of sticking (no slip, seen as a small vertical portion in the hysteresis plot) in the reduced model, which is negligibly small in the full model. This error does not visibly decrease even if, say, the six-state reduced model is changed to an eight-state reduced model. The error appears to be due to the error in approximating $G(\eta)$ above. However, for a large set of forcing histories, this error is slight. Overall, the minor loops are captured well; recall that the Bouc–Wen model does not capture such loops at all.

5. Conclusion

Our primary aim in this work was to systematically derive a reduced-order hysteresis model from a high-dimensional frictional system. Such an approach allows access to all internal states, and we hope that it yields insights into hysteretic phenomena, in general, and leads to more compact and easy to use hysteresis models in future work.

The model we have studied shows hysteresis, including major and minor loops. Such minor loops cannot be captured by the popular Bouc–Wen model. They can be captured by the Preisach model, but that model is very different from ours in both structure and function. Thus, our main contribution in this work lies in the reduction of a high-dimensional hysteretic system into a few explicit evolution equations. Our success has depended on a serendipitous \textit{ad hoc} approximation of the function $G(\eta)$ as well as on being able to solve a minimization problem involving this same function. We are not aware of similar approximations made elsewhere.

However, at this stage, our reduced model remains somewhat complicated. It has three fitted matrices, and the evolution rate involves evaluation of a slightly complicated matrix function which includes root finding of polynomials. Additionally, our numerical solution procedure leads to a stiff system. These difficulties suggest that smaller empirical models must still be sought, for efficient simulation of dynamical systems involving hysteretic elements. Yet, our results may provide insights that aid in such a search. Future work may yet simplify and improve our approach, making it more competitive for routine computational applications. In this way, our work opens up lines of new research into both the mathematics and physics of hysteresis.

Acknowledgements. We thank Atanu Mohanty and Tushar Saha for discussions.

Funding statement. The authors would like to thank the Department of Science and Technology, Government of India, for funding.

Appendix A. Derivation of frictional system equation and its linear complementarity problem form

We consider the system shown in figure 4. Suppose first that the blocks have masses $m_1, m_2, m_3, \ldots, m_N$. The kinetic energy (KE) and the potential energy (PE) of the system are as follow

$$KE = \frac{1}{2} \sum_{i=1}^{N} m_i x_i^2,$$

$$PE = \frac{1}{2} x^T K x,$$

\(^5\)See the lines following equation (3.16). If both $q = 0$ and $f = 0$, then $c = 0$, and hence $\bar{c}_i = 0$ for each $i$.\)
where \( x \) is an \( N \)-dimensional vector of generalized coordinates and \( K \) is a symmetric, positive definite matrix of size \( N \times N \). Lagrange’s equations of motion \([20]\) are

\[
M \ddot{x} + Kx = bf(t) + F,
\]

where \( M \) is a diagonal matrix whose \( i \)th element is \( m_i \); \( b \) is a column matrix whose \( i \)th element is \( b_i \) as indicated in \textit{figure 4}; and \( F \) is an \( N \)-dimensional vector whose \( i \)th element is \( F_i \) as indicated in \textit{figure 4}. Here, the system is massless, and the force vector \( F \) arises from Coulomb friction, and so we have

\[
\mu \text{ sgn}(\dot{x}) + Kx = b f(t).
\]

(A1)

In the above, \( \mu \) is an \( N \times N \) diagonal matrix with positive elements. Pre-multiplying equation (A1) by \( \mu^{-1} \) yields

\[
\text{sgn}(\dot{x}) + \mu^{-1}Kx = \mu^{-1}bf(t),
\]

which we write as

\[
-\tilde{F} + \tilde{K}x = \tilde{b} f(t),
\]

(A2)

where \( \mu^{-1}K = \tilde{K}, \mu^{-1}b = \tilde{b} \), and

\[
\text{sgn}(\dot{x}) = -\tilde{F}.
\]

(A3)

We now cast an incremental form of equation (A2) into LCP form. This way of solving friction problems is well known \([21]\) but is described here for completeness.

We introduce two new non-negative variables \( U \) and \( V \), defined elementwise as follows. Let \( U = -\dot{x} \) if \( \dot{x} < 0 \), and \( U = 0 \) otherwise; and \( V = \dot{x} \) if \( \dot{x} > 0 \) and \( V = 0 \) otherwise. We then have the complementarity relations

\[
U \geq 0, \quad V \geq 0, \quad U^T V = 0,
\]

and observe that

\[
\dot{x} = V - U.
\]

Next, we observe that \( |\tilde{F}| \leq 1 \). We introduce two more new non-negative variables \( S \) and \( R \), defined elementwise as follows. Let \( S = 1 + \tilde{F} \geq 0 \) and \( R = 1 - \tilde{F} \geq 0 \).

It follows that when (for any element) \( \dot{x} > 0 \), and \( V > 0 \), then \( \tilde{F} = -1 \) (see equation (A3)) or \( S = 0 \). Similarly, when \( \dot{x} < 0 \), and \( V = 0 \), then \( \tilde{F} = 1 \), or \( S > 0 \). Thus, \( V \) and \( S \) satisfy the complementarity relation \( V^T S = 0 \). It can be similarly seen that \( U^T R = 0 \).

We now note that

\[
\begin{align*}
S + R &= 2 \\
S - R &= 2\tilde{F}
\end{align*}
\]

(A4)

Considering a small time increment \( \Delta t \), equation (A2) becomes

\[
-\Delta \tilde{F} + \tilde{K} \Delta x = \tilde{b} \Delta f
\]

(A5)

Now, \( \Delta x = (V - U)\Delta t \) gives

\[
-\Delta \tilde{F} + \tilde{K} \Delta t (V - U) = \tilde{b} \Delta f,
\]

(A6)

where we write

\[
\Delta \tilde{F} = \frac{\Delta S - \Delta R}{2} = \frac{S_{k+1} - S_k - (R_{k+1} - R_k)}{2},
\]

(A7)

and consider writing

\[
\tilde{b} \Delta f = \tilde{b}(f_{k+1} - f_k).
\]

However, \( x \) does not appear explicitly in equation (A6). To avoid possible drift in \( x \), we retain \( x \) by writing

\[
\tilde{b} \Delta f = \tilde{b} f_{k+1} - \tilde{K} x_k + \frac{S_k - R_k}{2},
\]

(A8)
Finally, equations (A 6)–(A 8) are combined and rearranged to give

\[
\begin{bmatrix}
1 & 1 \\
-1/2 & 1/2
\end{bmatrix}
\begin{bmatrix}
S_{k+1} \\
R_{k+1}
\end{bmatrix} - \begin{bmatrix}
0 & 0 \\
-\tilde{K}\Delta t & \tilde{K}\Delta t
\end{bmatrix}
\begin{bmatrix}
V_{k+1} \\
U_{k+1}
\end{bmatrix} = \begin{bmatrix}
\tilde{b}\Delta t - \frac{2}{2} (S_k - R_k) \\
\tilde{b}\Delta t - \frac{2}{2} (S_k - R_k)
\end{bmatrix},
\]  

(A 9)

where the upper row in equation (A 9) is a direct restatement of equation (A 4).

The LCP considers given matrices \(\tilde{M}\) of size \(p \times p\) and \(\tilde{q}\) of size \(p \times 1\), and seeks vectors \(\tilde{w}\) and \(\tilde{z}\) which satisfy

\[
\tilde{w} - \tilde{M}\tilde{z} = \tilde{q},
\]

\[
\tilde{w}_i \geq 0; \quad \tilde{z}_i \geq 0;
\]

\[
\tilde{w}_i\tilde{z}_i = 0.
\]

Noting the complementarity relations between \(S\) and \(V\), and between \(R\) and \(U\), we find that equation (A 9) is an LCP with

\[
\tilde{M} = \begin{bmatrix}
1 & 1 \\
-1/2 & 1/2
\end{bmatrix}^{-1}
\begin{bmatrix}
0 & 0 \\
-\tilde{K}\Delta t & \tilde{K}\Delta t
\end{bmatrix}, \quad \tilde{q} = \begin{bmatrix}
1 & 1 \\
-1/2 & 1/2
\end{bmatrix}^{-1}
\begin{bmatrix}
\tilde{b}\Delta t - \frac{2}{2} (S_k - R_k) \\
\tilde{b}\Delta t - \frac{2}{2} (S_k - R_k)
\end{bmatrix}.
\]

The LCP can be numerically solved using Lemke’s algorithm, and our solutions were obtained using an implementation of the same given by Miranda & Fackler [17] in their ‘CompEcon’ toolbox. In our solutions, we start from an initial state, choose a time increment \(\Delta t\), and march forward in time to obtain a solution. Redoing with a smaller \(\Delta t\) and obtaining the same results indicates that the solution is sufficiently accurate for our purposes.

References


