Robust Extended Kalman Filter for Transient Tracking and Outlier Suppression

Mital A. Gandhi, Student Member, IEEE, and Lamine Mili, Member, IEEE

Abstract—A new filter is proposed that achieves reliable state estimation in nonlinear systems with multiple equilibrium points. The latter may exhibit strong nonlinearity and sudden transient behavior triggered by either system process noise, or observation noise, or outliers. Our filter is able to output the correct qualitative state over time quickly and reliably when the system dynamics experience switch-like transitions. It yields robust performance through two key features: (a) existence of observation redundancy in the filter equations and (b) application of generalized maximum likelihood (GM-) estimators when solving for the system states in an extended Kalman filter (EKF) methodology. Our filter, which we call the GM-EKF, is formulated in a batch-mode regression form to process the observations and predictions together. This observation redundancy allows the GM-EKF to track system transitions from one equilibrium point to another one much more reliably and in a faster way than the conventional EKF, even when the observation noise variances are much larger than the system process noise variances. As for the GM-estimator, it enables the filter to bound the influence of outliers while maintaining a high statistical efficiency when the system is resident around a stable equilibrium point. Results are presented from simulations applying the EKF and GM-EKF on the Langevin model, which is commonly used to model climate transitions. These simulations reveal the improvements in estimation efficiency, robustness, and reduced time delay when tracking the transitions.

Index Terms—Robust estimation, GM-Kalman filter, Nonlinear filtering.

I. INTRODUCTION

The focus of this paper is state estimation in nonlinear systems with multiple equilibrium points, whereby the state may suddenly shift from one point to another. Such transitions may be driven by the system process noise or observation noise, both of which are often assumed to follow a Gaussian probability distribution function (PDF). However, the assumed noise model is only an approximate one and the two types of noises may actually follow thick-tailed, non-Gaussian PDFs, inducing innovation and observation outliers. These outliers arise naturally in many ways in practice, such as hardware discontinuities in control systems and faults in the sensors of a system [1], [2], and may easily lead to erroneous estimates. Thus, a robust filter is desired that is able to output the correct qualitative state over time quickly and reliably in the presence of both strong, nonlinear, “switch-like” transitions [3]–[5] and outliers.

The need for such a filter is growing in various fields, including meteorology, physical oceanography, and paleoclimatology [6], [7]. The vast majority of numerical techniques in these areas are based on variants of the least-squares estimator and the Kalman filter (KF) [8]. Regarding the KF, the system’s dynamics are represented by state space dynamic and observation models that account for the system process and observation noises, respectively. It provides a maximum likelihood estimate assuming a quadratic performance criterion and Gaussian PDFs for the noises. However, it is designed for linear models, and therefore, is not effective in reliably tracking the nonlinear state transitions from one equilibrium point to another.

Various other methods have been considered in these areas [4], [5], [8], including the least squares variational method (LSV), the extended Kalman filter (EKF), and the interactive Kalman filter. The LSV is actually an example from a class of methods developed in statistical physics, in which variational principles have been formulated as an optimization problem to determine the time-sequence of states [4], [5]. However, this method is unable to track the transient well because it is generally developed for problems with weak noise [4], which means that the probability of large deviations from an operating point of a nonlinear system is relatively low. In other words, sudden shifts from one equilibrium point to another are assumed to be unlikely, an assumption that is violated in many practical problems.

One popular method for state estimation in nonlinear systems is the extended Kalman filter (EKF). This filter is able to use a nonlinear model directly in the prediction stage, in which linearization and discretization are performed around the previous filtered state estimate to obtain a set of discretized, linear perturbation equations. The classical KF equations are then used as a basis for the EKF filtering equations. Similar to the LSV, the EKF is also developed for problems with weak noise; therefore, it also does not track the system transitions well in the presence of Gaussian noise, and yields worse solutions or diverges completely in the presence of outliers [4], [9]. An example of the EKF’s performance under weak and strong noise can be found in the work of Picard [10], which observes conditions under which the EKF may become or remain small asymptotically.

Burger [3] proposed the method known as the interactive Kalman filter, in which several error models are simultaneously used to solve for the state estimate. In this approach, several PDFs are explicitly assumed for the noise processes, and the system’s state estimate is determined by the noise PDF that most closely represents the current system behavior. One disadvantage of this method is the need for a priori definition...
of specific noise models. In fact, the technique may easily lead to inaccurate results and increased computational costs if one uses too many noise models to represent the system, or on the other hand, under-represents the system with an insufficient number of noise regimes.

Yet other methods from the literature that may be used to track state transitions include the particle filter, particle Kalman filter, path integration, the ensemble Kalman filter and its second-order variant, and the singular extended interpolated filter [4], [8], [11], [12]. However, these techniques are not attractive as they are often very complicated and computationally very intensive, requiring the use of Monte-Carlo principles in determining how the system’s PDF evolves over time [13].

In contrast to these methods, we propose in this paper a filter to reliably track the states of a nonlinear dynamic system. Our filter is simple and robust against outliers also. Its design framework consists of three key steps, which include (a) creating a redundant observation vector, (b) performing robust prewhitening, and (c) robustly estimating the state vector. This framework allows us to employ one of many robust estimators within the extended Kalman filter methodology, so long as that estimator’s covariance matrix can be calculated. In this paper, we utilize the generalized maximum likelihood-type (GM-) estimator, yielding our proposed filter known as the GM-EKF.

We now discuss the three steps in more detail. In the first step, we convert the classical recursive approach into a batch-mode regression form so that the observations may be processed simultaneously. Note that observation redundancy is required for an estimator to be capable of suppressing the outliers and tracking the transitions well, and can be achieved in practice by simply placing more sensors in the system. The second step consists of applying a prewhitening procedure that utilizes a robust estimator of location and covariance such as the Projection Statistics (PS) [14]–[17] or the minimum covariance determinant (MCD) [18]. The prewhitening procedure robustly uncorrelates the noise when outliers are present in the predictions and observations. In the third step, the unconstrained nonlinear optimization in the GM-estimator is solved using the Iteratively Re-weighted Least Squares (IRLS) algorithm. Finally, the influence function (IF) of the GM-estimator is employed to derive the asymptotic state estimation error covariance matrix of the GM-EKF [19]–[21]. Recall that while the IF of an estimator is a measure of its sensitivity to infinitesimal contamination at a given distribution, its covariance matrix is equal to that of the estimator at that distribution [22], [23].

This paper is organized as follows. Section II contains background material on the nature of nonlinear systems with multiple equilibrium points and the Langevin model. Section V presents the EKF and demonstrates its limitations when applied to climate transition tracking. The GM-EKF is developed in Section III. The influence function of the GM-estimator for nonlinear models is derived in Section IV. In Section V-B, the GM-EKF is shown to track the state transitions in the Langevin equation. Conclusions are drawn in Section VI.

II. BACKGROUND TOPICS

A. Nonlinear Systems with Multiple Equilibrium Points

In this paper, we investigate a class of nonlinear models that contain multiple equilibrium points. As an example, let us consider the general system depicted in Figure 1 and the associated contour plot in Figure 2. We see that this system contains three stable equilibrium points, with uneven saddle surfaces that act as boundaries between three regimes of the system. It has been observed that such systems often have small fluctuations around an equilibrium point, followed by a sudden shift in the system state to another equilibrium point. Thus, the system will reside in one of the basins of attraction for a certain amount of time, followed by a transition that is caused by noise with large enough energy to move it from one basin to another. Of course, the amount of energy required depends on how deep the basin is compared to the energy of the lowest saddle point along the boundary. What is desired then is a method that is able to accurately and quickly determine the system’s states as it moves between these basins.

B. Review of the Extended Kalman Filter

We now review the extended Kalman filter that is popularly applied for state estimation in such systems. The dynamics of a general nonlinear system, assuming a vector state variable, is
represented in the EKF by the following equations [24]–[26]:

\[ \dot{x}_t = f(x_t) + w_t + u_t, \quad (1) \]
\[ z_t = h(x_t) + e_t, \quad (2) \]

where \( f(x_t) \) and \( h(x_t) \) are assumed to be continuous and continuously differentiable with respect to all elements of the state vector, \( x_t \). The terms \( w_t \) and \( e_t \) in (1) and (2) are the system process noise and observation noise processes, respectively, and \( u_t \) is the control input vector. Note that we do not have to explicitly model this external forcing term as the energy from \( w_t \) may also drive the system. Let us denote by \( W_t \) and \( R_t \) the noise covariance matrices in the continuous time. To solve for the state estimates, the model must first be linearized around the previous corrected estimate. Applying a first-order Taylor series expansion to (1) and assuming the higher order terms are negligible, we get

\[ \delta \dot{x}_t = F_x(x^*_t) \delta x_t + w_t, \quad (3) \]

where \( F_x \) is the Jacobian matrix defined as

\[ F_x = \frac{\partial f(x_t)}{\partial x_t} \bigg|_{x_t=x^*_t}, \quad (4) \]

and \( x^*_t \) is the nominal state vector. Let \( x_0 \) denote the initial condition. For successive iterations, the corrected estimate \( \hat{x}_{k-1|k-1} \) is assigned as the nominal value at time \( k \). Similarly, linearizing the nonlinear observation equation yields the following perturbation equation:

\[ \delta z_t = H_x(x^*_t) \delta x_t + e_t, \quad (5) \]

where \( H_x \) is a Jacobian matrix given by

\[ H_x = \frac{\partial h(x_t)}{\partial x_t} \bigg|_{x_t=x^*_t}. \quad (6) \]

Using these perturbation equations and assuming zero-order hold on any inputs and continuous integration of the noise, the nonlinear model can be discretized and put into the following form [25], [27]:

\[ x_k = F_d x_{k-1} + w_k, \quad (7) \]
\[ z_k = H_d x_k + e_k, \quad (8) \]

where

\[ F_d = e^{F_x T}, \quad (9) \]
\[ H_d = H_x. \quad (10) \]

and \( T_s \) is the time step. A detailed derivation that shows the homogeneous and particular solutions of (3) can be found in various texts [25], [27]–[29].

Because this is a discrete system with linear matrices, the solution can be written similarly to the traditional linear Kalman filter recursions, as follows [25]:

\[ \hat{x}_{k|k-1} = F_d \hat{x}_{k-1|k-1} + \int_{k-1}^{k} f(\hat{x}_{k-1|k-1}) \, dt, \quad (11) \]
\[ \Sigma_{k|k-1} = F_d \Sigma_{k-1|k-1} F_d^T + W_k, \quad (12) \]
\[ K_k = \Sigma_{k|k-1} H_d^T \left( H_d \Sigma_{k|k-1} H_d^T + R_k \right)^{-1}, \quad (13) \]
\[ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k [z_k - h(\hat{x}_{k|k-1})], \quad (14) \]
\[ \Sigma_{k|k} = \Sigma_{k|k-1} - K_k H_d \Sigma_{k|k-1}. \quad (15) \]

C. Challenges in State Estimation

Now, we discuss several challenges that may arise when the EKF method is applied to track the states of a nonlinear system with multiple equilibrium points. First, the EKF relies heavily on accurate observations to properly track state transitions. To understand this, note that the system process noise is not directly included in the state prediction equation in (11), and that its effects are actually incorporated via its covariance in (12) and (15). Effectively, the observation noise covariance needs to be sufficiently lower than the system process noise covariance for the filter to appropriately rely on the observations and properly sense the transitions.

On the other hand, when the observation noise covariance is sufficiently larger than the system process noise covariance, the EKF may miss a transition by ignoring valid observations while incorrectly relying on the predictions, which do not detect the transitions in the first place. Moreover, the filter may unduly output a state transition in its estimate as a result of an innovation outlier in the predictions.

Finally, the filter’s dependence on the observations makes it very sensitive to observation outliers as well. Indeed, the filter may output a transition due to an outlying observation, ignoring a good prediction even when the system has not actually shifted between the basins of attraction.

III. DEVELOPMENT OF THE GM-EKF

These drawbacks are demonstrated by simulating a simple climate model in Section V-B. In particular, it will be seen that the EKF leads to inaccurate estimates very easily in the presence of outliers, because just like the Kalman filter, the recursive algorithm solves a least squares estimator with no resistance to outliers. Furthermore, the standard formulation of the EKF degrades in many ways even when no outliers are present. Here, we will develop a new robust filter with the following key advantages:

- Redundancy in the observations to overcome the sensitivity to the observation variance and overconfidence in the predictions
- Use of robust estimates of variance along with a reliable iterative algorithm, making the filter resistant against outliers. The latter in this context may be thought of as excessively large fluctuations that may inaccurately cause a shift in the state estimates, whereas the true system has not shifted. In our application, we use the Projection Statistics (PS) and the Iteratively Re-weighted Least Squares (IRLS) algorithms to achieve our goals.
- Smooth and gradual down-weighting function that helps maintain high statistical efficiency.

The development of the GM-EKF in this paper is actually an extension of the robust filter developed for linear systems, known as the GM-KF and derived in detail in [19]. We now describe the key steps of the GM-EKF method. Let the state transition and observations equations be given by (1) and (2) in the continuous time, respectively. The first step of the GM-EKF is to formulate the problem in a batch-mode regression form and develop a redundant observation vector. For this, the model is first discretized and linearized yielding the system...
equations in discrete time expressed in (7) and (8). Then, we predict the state of the system using the same equation from the EKF, given by (11). One may use one of several numerical integration methods to compute the integral in this equation, such as the first-order Euler and fourth-order Runge-Kutta methods. Then, we combine these predictions with the observations to obtain the batch regression form, as follows:

\[
\begin{bmatrix}
  z_k \\
  \hat{x}_k[k|k-1]
\end{bmatrix} =
\begin{bmatrix}
  H_d & I
\end{bmatrix}_k
\begin{bmatrix}
  \hat{x}_k \\
  e_k
\end{bmatrix} +
\begin{bmatrix}
  \delta_{k|k-1}
\end{bmatrix},
\]

where \(\delta_{k|k-1}\) is the error between the true state and its prediction, yielding

\[
\tilde{z}_k = \hat{H}x_k + \tilde{e}_k.
\]

The covariance matrix \(\tilde{R}_k\) of the error \(\tilde{e}_k\) is given by

\[
\tilde{R}_k =
\begin{bmatrix}
  R_k & 0 \\
  0 & \Sigma_{k|k-1}
\end{bmatrix},
\]

with \(R_k\) being the noise covariance matrix of \(e_k\) and \(\Sigma_{k|k-1}\) the propagated filter error covariance matrix given by (12).

The second step of the GM-EKF is to uncorrelate the data. However, instead of using the influence function of the GM-EKF, we use a robust data pre-whitening procedure in our proposed filter. In this procedure, we must first identify the outliers using a robust outlier detection method. In this paper, we use the Projection Statistics (PS) algorithm as described in [14], [30]. Then, we compute the weights \(\omega_i\) for the elements of the vector \(\tilde{z}_k\) using the PS values, as follows:

\[
\omega_i = \min \left(1, \frac{d^2}{PS_i^2}\right).
\]

Next, the outliers are down-weighted by applying \(\omega_i\) to the elements of \(\tilde{z}_k\). We continue the robust pre-whitening procedure as follows. Using either upper diagonal factorization or Cholesky decomposition, we obtain the matrix \(S_k\) such that \(R_k = S_kS_k^T\). Equivalently, we may use the square-root method to obtain \(\sqrt{R_k}\) such that \(\sqrt{R_k} = \sqrt{R_k}\sqrt{R_k}\). Finally, we multiply the linear regression model \(\tilde{z}_k = \hat{H}x_k + \tilde{e}_k\) on the left-hand side by \(S_k\) or \(\sqrt{R_k}\) to perform pre-whitening; for example,

\[
(\sqrt{R_k})^{-\frac{1}{2}}\tilde{z}_k = (\sqrt{R_k})^{-\frac{1}{2}}\hat{H}x_k + (\sqrt{R_k})^{-\frac{1}{2}}\tilde{e}_k,
\]

yielding the final form of the regression as

\[
y_k = A_kx_k + \eta_k.
\]

The third and last step of the GM-EKF is to solve for the state estimates iteratively using the GM-estimator with the IRLS algorithm. Formally, the GM-estimator is defined as that which minimizes the objective function

\[
J(x) = \sum_{i=1}^{m} \omega_i^2 \rho \left( \frac{r_i}{s\omega_i} \right),
\]

where \(\rho(\cdot)\) represents a nonlinear function of standardized residuals \(r_i\), defined as

\[
r_i = y_i - \hat{a}_i^T x.
\]

The robust scale \(s\) in (22) is the median absolute deviation, defined as \(s = 1.483\ \text{median}_{i} |r_i|\). In this work, we use the Huber \(\rho\)-function, given by

\[
\rho \left( \frac{r_i}{s\omega_i} \right) = \begin{cases}
\frac{1}{2} \left( \frac{r_i}{s\omega_i} \right)^2, & \text{for } \frac{r_i}{s\omega_i} < c \\
\frac{c}{2} \left( \frac{r_i}{s\omega_i} \right)^2 - \frac{c^2}{2}, & \text{elsewhere},
\end{cases}
\]

where \(c = 1.5\) is a cutoff threshold that yields good statistical efficiency at the Gaussian distribution without increasing the bias too much under contamination [22], [31]. In other words, this function provides high efficiency at Gaussian noise but a bounded and continuous influence function for outliers by reducing to \(L_2\)-norm for small residuals and to \(L_1\)-norm for large residuals.

The GM-estimator is obtained by setting the partial derivatives of the objective function to zero, yielding

\[
\frac{\partial J(x)}{\partial x} = \sum_{i=1}^{m} \omega_i a_i^T \psi \left( \frac{r_i}{s\omega_i} \right) = 0,
\]

where

\[
\psi \left( \frac{r_i}{s\omega_i} \right) = \frac{\partial \rho \left( \frac{r_i}{s\omega_i} \right)}{\partial \left( \frac{r_i}{s\omega_i} \right)}.
\]

Solving this system of nonlinear equations using the IRLS algorithm then yields

\[
\hat{x}_{k|k}^{T+1} = \left( A_k^T Q^{(\nu)} A_k \right)^{-1} A_k^T Q^{(\nu)} y_k,
\]

where the weight matrix \(Q\) is given by

\[
Q = \text{diag} \left\{ q \left( \frac{r_i}{s\omega_i} \right) \right\},
\]

with

\[
q \left( \frac{r_i}{s\omega_i} \right) = \psi \left( \frac{r_i}{s\omega_i} \right) \frac{r_i}{s\omega_i},
\]

and \(\omega_i\) is given by (19).

IV. INFLUENCE FUNCTIONS OF GM-ESTIMATORS FOR NONLINEAR REGRESSION MODELS

Linearization of the nonlinear model around the nominal values of the estimate enabled a simple formulation of the GM-EKF. The filter error covariance matrix \(\Sigma_{k-1|k-1}\) is a key component of this filter, as it is required to compute \(\tilde{R}_k\) in the pre-whitening procedure and also influences the state estimation via (27). For the EKF, \(\Sigma_{k-1|k-1}\) may be computed using the known covariance matrices \(R_k\) and \(W_k\) following (15). For the robust GM-EKF though, it needs to be developed using the following equation that relates the influence function of the estimator to the asymptotic covariance matrix [19]:

\[
\Sigma = E[\text{IF}^{\text{IF}^T}].
\]

However, instead of using the influence function of the GM-estimator along with (30) as described in [19], one needs to derive the IF corresponding to the particular nonlinear model under consideration. Next, we derive the general form of
this influence function. Let the nonlinear regression model be expressed as

\[ y = \varphi(a, x) + \eta, \quad (31) \]

where \( y \) are the observations, \( a \) are the explanatory variables, \( x \) is the parameter vector, \( \eta \) is the observation noise, and \( \varphi \) is a vector-valued nonlinear regression function with respect to \( a \) and/or \( x \). Recall that \( y \) and \( \eta \) in this model are i.i.d. vectors following Gaussian distributions. Assuming the function \( \varphi \) is deterministic and independent of the residuals, the cumulative probability distribution function of the residual error vector \( r \), expressed as

\[ r = y - \varphi(a, \hat{x}), \quad (32) \]

is denoted by \( \Phi(r) \). For this model, the GM-estimator in regression provides an estimate for \( x \) by processing the redundant observation vector \( y \) and solving the implicit equation given by

\[ \sum_{i=1}^{m} \lambda_i(r, a, x) = 0, \quad (33) \]

where

\[ \lambda_i(r, a, x) = \omega_i \frac{\partial \varphi_i(a, x)}{\partial x} \psi(r_{s_i}), \quad (34) \]

and \( r_{s_i} = r/s \omega_i \). Given the empirical cumulative probability distribution function \( F_m \), the functional form of the estimator, where \( x \) is replaced by \( T \), is given by the vector-valued functional

\[ \int \lambda(r, a, T) dF_m = 0. \quad (35) \]

Asymptotically, \( F_m \to G \) by virtue of the Glivenko-Cantelli Theorem [32] and (35) becomes

\[ \int \lambda(r, a, T(G)) dG = 0. \quad (36) \]

Following Neugebauer [20] and Thomas [21], we begin the derivation of the asymptotic influence function, given by

\[ \text{IF}(r, a; \Phi) = \left. \frac{\partial T(G)}{\partial \epsilon} \right|_{\epsilon=0} = \lim_{\epsilon \to 0} \frac{T((1-\epsilon)\Phi + \epsilon H) - T(\Phi)}{\epsilon}, \quad (37) \]

by substituting \( G = (1-\epsilon)\Phi + \epsilon H \) into (36), yielding

\[ 0 = \int \lambda(r, a, T(G)) dG \quad (38) \]

\[ = \int \lambda(r, a, T(G)) d\Phi + \epsilon \int \lambda(r, a, T(G)) d(H - \Phi). \quad (39) \]

Differentiating with respect to \( \epsilon \) and applying the chain rule yields

\[ \frac{\partial}{\partial \epsilon} \int \lambda(r, a, T(G)) d\Phi + \int \lambda(r, a, T(G)) d(H - \Phi) \quad (40) \]

Next, recall the following interchangeability of differentiation and integration:

\[ \int_S \frac{df(\epsilon, x)}{\partial \epsilon} \mu(dx) = \int_S f'(\epsilon, x) \mu(dx), \quad (40) \]

assuming that \( f \) is continuous and measurable and \( f' \) is measurable on \( S \). Using this result and given \( H = \Delta_r \), where \( \Delta_r \) is the probability mass at \( r \), the above equation reduces to

\[ \int \frac{\partial}{\partial \epsilon} \lambda(r, a, T(G)) d\Phi + \int \lambda(r, a, T(G)) d(\Delta_r - \Phi) = 0. \quad (41) \]

Evaluating this expression at \( \epsilon = 0 \) makes the last term equal to zero. It follows that

\[ \int \frac{\partial}{\partial \epsilon} \left[ \lambda(r, a, T(G)) \right]_{\epsilon=0} d\Phi + \int \lambda(r, a, T(G)) d\Delta_r = \int \lambda(r, a, T(G)) d\Phi. \quad (42) \]

Assuming Fisher consistency at \( \Phi \), the right-hand side of (42) becomes zero. Then, using the sifting property of \( \Delta_r \), we obtain

\[ \int \left[ \frac{\partial}{\partial \epsilon} \lambda(r, a, x, \Phi) \right]_{\epsilon=0} d\Phi + \lambda(r, a, T(\Phi)) = 0. \quad (43) \]

This results in the following expression for the influence function:

\[ \text{IF}(r, a; \Phi) = \left. \frac{\partial T(G)}{\partial \epsilon} \right|_{\epsilon=0} = - \left[ \int \frac{\partial}{\partial \epsilon} \lambda(r, a, x, \Phi) \right]_{\epsilon=0}^{-1} \lambda(r, a, T(\Phi)). \quad (44) \]

Substituting (34) into (44) yields

\[ \text{IF}(r, a; \Phi) = - \left[ \int \frac{\partial}{\partial \epsilon} \lambda(r, a, x, \Phi) \right]_{\epsilon=0}^{-1} \lambda(r, a, T(\Phi)). \quad (45) \]

Deriving \( \lambda(\cdot) \) with respect to \( x \), and assuming that \( \omega \) and \( s \) are independent of \( x \) [21], we obtain

\[ \frac{\partial \lambda(y, a, x)}{\partial x} = \omega \left[ \frac{\partial \psi(r_{s_i})}{\partial x} \right] \left[ \frac{\partial \varphi(y, x)}{\partial x} \right]^T + \omega \psi(r_{s_i}) \frac{\partial^2 \varphi(y, x)}{\partial x_1 \partial x_j}, \quad (46) \]

where \( \frac{\partial^2 \varphi(y, x)}{\partial x_1 \partial x_j} \) is the Hessian matrix of \( \varphi(y, x) \). Applying the chain rule to the derivative of \( \psi(\cdot) \) yields

\[ \frac{\partial \lambda(y, a, x)}{\partial x} = - \psi'(r_{s_i}) \left[ \frac{\partial \varphi(y, x)}{\partial x} \right] \left[ \frac{\partial \varphi(y, x)}{\partial x} \right]^T + \omega \psi(r_{s_i}) \left[ \frac{\partial^2 \varphi(y, x)}{\partial x_1 \partial x_j} \right], \quad (47) \]

Finally, substituting (34) and (47) into (45) yields the following as the influence function of the GM-estimator for nonlinear regression:

\[ \text{IF}(r, a; \Phi) = - \left[ \int \frac{\partial}{\partial \epsilon} \lambda(r, a, x, \Phi) \right]_{\epsilon=0}^{-1} \lambda(r, a, T(\Phi)). \quad (45) \]
latter two are stable. In other words, starting with any points other than the origin, the system will approach and settle at either \( x = 1 \) or \( x = -1 \). With small forcing or noise terms, the system will fluctuate around one of these equilibrium points. With a large enough fluctuation or constant external forcing term or inputs driving the system in the correct direction, the system will shift from one basin of attraction to another [3], [8], which is termed a “transition” in the system.

In practice, similar models are quite frequently used as a simple representation of climate system exhibiting transient behavior [4], [8], [35]-[38]. The function \( U(x_t) \) given by (50) can be interpreted as the climate potential, with \( x_t \) representing the average surface temperature of the earth at time \( t \). The right minimum can be viewed as the present climate state while the other equilibrium represents ice ages, for example [4], [37]. With small values of \( \kappa \), the state dynamics consists of small fluctuations around an equilibrium point for a long-term with a large fluctuation leading to transitions occasionally. However, such behavior is not very applicable to long-time climate dynamics that are analyzed on geological time scales, in which transitions occur frequently [4], [8].

The evolution of the probability distribution of the system in (51) can also be derived explicitly. In general, the PDF \( F(x_t, t) \) of the state \( x_t \) described by the Ito stochastic differential equation, given by

\[
dx_t = f(x_t, t)dt + g(x_t, t)db_t,
\]

(52)
can be obtained through the Fokker-Planck equation, where \( g(x_t, t) \) is an arbitrary function, \( b_t \) is a Wiener process, and \( f(x_t, t) \) specifies the system dynamics. Also known as the forward Kolmogorov equation, it is expressed as

\[
\frac{\partial F(x_t, t)}{\partial t} = -\nabla \cdot (f(x_t, t)F(x_t, t)) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} (R_{ij})_{ij} F(x_t, t),
\]

(53)
where \( R_{ij} = g(x_t, t)g^T(x_t, t) \) [8], [13]. For the climate model described in (51), the Fokker-Planck equation is expressed as

\[
\frac{\partial F(x_t, t)}{\partial t} = \frac{\partial}{\partial x_t} \left[ 4x_t (x_t^2 - 1) F(x_t, t) \right] + \frac{1}{2} \kappa^2 \frac{x_t^2}{\partial x_t^2} F(x_t, t),
\]

(54)
which yields the steady state PDF solution [39] of the following form:

\[
F(x_t, t) \propto e^{-2U(x_t,t)/\kappa^2},
\]

(55)
where \( U(x_t, t) \) is given by (50). Though we may estimate the PDF directly, it should be noted that Monte-Carlo type solutions such as the particle filter that directly track the evolution of the PDF over time are generally computationally intensive. Therefore, we have focused our work on robustifying the simple EKF method.

---

**Fig. 4.** The double-well system dynamic equation, \( f(x) = -U'(x) \), shown with its 3 equilibrium points.

**Fig. 3.** The double-well potential function, \( U(x) \), shown with its 3 equilibrium points.

**V. THE EKF AND GM-EKF FOR CLIMATE TRANSITION TRACKING**

**A. The Langevin Model**

Now that we have developed the GM-EKF, let us apply it to a simple model with two equilibrium points. In particular, the system model we consider in this paper is given by the dynamic equation expressed as [8]

\[
\dot{x}_t = f(x_t) = -4x_t(x_t^2 - 1),
\]

(49)
where the right-hand side can be viewed as the negative gradient of a potential function, \( U(x_t) \), given by

\[
U(x_t) = x_t^2 - 2x_t^2.
\]

(50)
With a stochastic forcing term, the dynamic equation becomes

\[
\dot{x}_t = 4x_t - 4x_t^3 + \kappa n_t,
\]

(51)
which is termed the Langevin model when \( n_t \) is a white-noise process.

The double-well potential function represented by (50) is shown in Figure 3, with the negative of the gradient, \( f(x) \), shown in Figure 4. It is clear that without any external forcing term or noise, this system has three equilibrium points at \( x = 0 \), \( x = 1 \), and \( x = -1 \), of which the former is unstable and
TABLE I
TEST CASES FOR EVALUATING THE EKF AND GM-EKF ON A DOUBLE-WELL SYSTEM

<table>
<thead>
<tr>
<th>Transition</th>
<th>Length (samples)</th>
<th>$\sigma^2_z$</th>
<th>$\sigma^2_x$</th>
<th>Sampling Frequency (Hz)</th>
<th>Sampling Period (s)</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1: Equilibrium conditions (no transitions)</td>
<td>-</td>
<td>0.01</td>
<td>0.01</td>
<td>4.00</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>Case 2: Effects of low sampling frequency</td>
<td>-</td>
<td>0.01</td>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
</tr>
<tr>
<td>Case 3: Effects of larger $\sigma^2_z$ compared to $\sigma^2_x$</td>
<td>-</td>
<td>0.01</td>
<td>0.01</td>
<td>4.00</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>Case 4: Rapid transitions under large values of $\sigma^2_z$</td>
<td>5</td>
<td>0.01</td>
<td>0.06</td>
<td>4.00</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>Case 5: Effects of observation outliers (no actual state transition)</td>
<td>-</td>
<td>0.01</td>
<td>0.01</td>
<td>4.00</td>
<td>0.25</td>
<td>0.75</td>
</tr>
</tbody>
</table>

B. Application of the EKF and GM-EKF to the Langevin Equation

We now apply the EKF and GM-EKF to the Langevin model and compare their performances in tracking state transitions. In particular, the test cases we have simulated are given in Table I. These tests are designed to evaluate the filters in the following situations:

1) A system resident around one of its stable equilibrium points, with no transitions;
2) Low frequency of observations during processing;
3) System perturbed by observation noise that has much larger covariance compared to system process;
4) System perturbed by outliers;
5) State transitions in the system occurring very rapidly.

Figure 5 shows the state estimation results for case 1, which can be considered in a general sense as evaluating the statistical efficiency of the filter (i.e. how well the filter works at the Gaussian distribution). Note that Gaussian ambient noise is assumed in Figure 5 and subsequent results in this section, with values of $\sigma^2_z = 0.01$, $\sigma^2_x = 0.01$, and $\kappa = 0.75$ unless otherwise noted. Also note that $\sigma^2_z$ and $\sigma^2_x$ are the observation and system process noise variances, respectively. Furthermore, note that the observation matrix for the GM-EKF is given by $H_d = [1, 1]^T$, i.e. it contains one redundant observation. As expected, the EKF estimates the states with good accuracy as seen by the very low mean squared error (MSE) values depicted in the bottom frame of the figure. The GM-EKF also performs equally well in this case, demonstrating that the filter maintains the high statistical efficiency and performance in the classical Gaussian noise situation.

Interesting limitations of the EKF and the advantages of the GM-EKF begin to surface when we simulate the system with transitions and observe how quickly, if at all, the filter tracks the system’s shift in equilibrium points. First of all, we have seen in simulations that when the sampling frequency is 4 Hz, i.e. each observation is collected in the sampling, the EKF detects the state shift with a time delay of 2 seconds. But, when the sampling frequency is 1.0 Hz so that every third sample of the observations is processed, the EKF detects the transition after a delay of 3 seconds whereas the GM-EKF is detects the transition as soon as observations are available, as seen in Figure 6. Reducing the frequency further to 0.50 Hz causes the EKF to oscillate for about 5 seconds until the transition is tracked properly while the GM-EKF continues to yield an accurate result immediately. Thus, a lower sampling frequency affects the EKF negatively the GM-EKF maintains its stable and fast performance by using just one redundant observation.

Next, we evaluate the filters’ sensitivity to the magnitudes of the noise variances. In particular, we show that if the observation noise variances are large enough in comparison to the system process noise and filter estimation error variances, the EKF will completely rely on the predictions, missing one or more transitions completely. As shown in Figure 7, if the observation noise variance in the model is set to $\sigma^2_z = 0.07$ instead of $\sigma^2_z = 0.01$, the EKF detects the state change after a delay of over 18 samples, which is over 4 seconds at the sampling frequency of 4 Hz in this experiment. Furthermore,
if the observation noise variance is increased to $\sigma_z^2 = 0.08$, the EKF does not track the state transition for at least 40 samples after the system transition occurs, i.e. at least 10 seconds, as seen in Figure 8. In comparison, one can see that the GM-EKF delivers a stable and fast result in both of these cases.

Qualitatively, this type of behavior can be explained by an overconfidence in the predictions by the EKF in comparison to the observations [4], [8], [13]. In particular, it turns out that the filter error variance approaches a steady state value of 0.05, or 22% error about the mean. Thus, when the observation noise variance is equal to 0.08, or 28% error about the mean, the observations no longer affect the system and do not force the model into a different equilibrium. In the general case, the EKF will exhibit this type of behavior due to the way the Kalman filter gain matrix and estimates are computed in (13) and (14). In fact, if the observations are accurate enough to make this filter gain larger than 0.50, then the filter is expected to correctly track the system transition. As indicated by Miller [13], the EKF can be made to follow the state transitions by processing the observations quickly enough such that the filter error variance does not reach its steady state value and the observations can “nudge” the estimation result to the accurate basin, which is what happens in Figure 7. However, if the observation noise is sufficiently greater than the system process noise, the filter gain may not become large enough even over a long period, causing the EKF to be too confident about the predictions and provide inaccurate state estimates for a long time or maybe even indefinitely as in Figure 8.

On the other hand, the GM-EKF performs well by doing exactly the opposite. Indeed, the GM-EKF does not depend completely on the predictions or the observations. Instead, it leverages the observation redundancy and the underlying GM-estimator to estimate the state more robustly. The latter relies on robust estimators of scale and covariance and down-weights any non-conforming data, including the predictions. In other words, though no single observation is accurate enough to make the filter gain larger than 0.50, the GM-EKF computes a gain large enough to output the correct state estimate using the observations and predictions together, and therefore, correctly senses the transition immediately.

Next, we observe scenarios in which the EKF tracks rapid transitions with an unduly time delay or is totally incapable of tracking rapid transitions when the observation noise variance is too large. As seen in Figure 9, the system shifts from an equilibrium point around -1 to another around +1, and back again within 5 samples. With $\sigma_z^2 = 0.05$, the observations are not accurate enough to track the transition immediately, but nevertheless, the filter does make the shift after a finite delay. But, when $\sigma_z^2 = 0.06$, the EKF does not detect the transition, as shown in Figure 10. In this case, the latter does not last long enough either to allow a sequence of observations to be processed, which could have eventually forced the filter to switch to another equilibrium point. Therefore, the EKF completely fails to follow the system’s state. Again, the GM-EKF performs yields accurate results with no lag in both of these cases.

Finally, we look at a situation in which several observation samples are outliers. Figure 11 depicts an experiment in which sequential outliers are induced from samples $n = 45$ to $n = 55$. In this case, the EKF filter inaccurately outputs multiple state transitions because it is unable to distinguish between good and bad observations whereas the GM-EKF filter appropriately suppresses the outlying observation at each
is shown in [40] that the maximum breakdown point of any estimator satisfies the property of general position, which requires every set of \( n \) row vectors of the matrix \( H \) in the linear regression model, \( z = Hx + e \), to be linearly independent [40]. After performing extensive simulations with the system containing several redundant observations and a varying number of outliers, we observe that the finite-sample breakdown point of the GM-EKF is no larger than 25%. Table II contains some results from these simulations showing how many outliers the filter is able to suppress versus the maximum possible breakdown point given by (57). It should be noted that the batch-mode form given by (17) provides a total of \( m_k = m + n \) observations in (57), not just \( m \).

VI. CONCLUSIONS

In this paper, a batch-mode GM-extended Kalman filter is proposed that tracks well state transitions in nonlinear systems in Gaussian noise and under contamination. A new error covariance matrix for the filter is developed using the influence function. We see that the GM-EKF performs well for the Langevin model at the standard Gaussian distribution, indicating a high efficiency level. Meanwhile, the GM-EKF also tracks the transitions well under other conditions, such as when the observation frequency is low, the observation noise variances are high, or outliers are present.

Many aspects of the GM-EKF are open to further research. The \( \rho \)-function can be modified to obtain better convergence rate, robustness, and efficiency properties in the filter. For example, replacing the Huber \( \rho \)-function with the strictly convex logistic function allows one to iterate by means of the Newton method instead of the IRLS algorithm. One may also use a different robust estimator of location and covariance instead of the Projection Statistics method applied in this work. As indicated earlier, this framework allows for the development of completely new filters also by replacing the GM-estimator with another regression estimator of choice.

REFERENCES

TABLE II
GM-EKF BREAKDOWN POINT IN THE PRESENCE OF OUTLIERS

<table>
<thead>
<tr>
<th>Total # of Observations</th>
<th># of Redundant Observations</th>
<th>Maximum # of Outliers that can Suppressed</th>
<th>Maximum Possible Breakdown Point</th>
<th>Actual # of Outliers Suppressed</th>
<th>Experimental Breakdown Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m + n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>33%</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>25%</td>
<td>1</td>
<td>25%</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>40%</td>
<td>2</td>
<td>40%</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>3</td>
<td>33%</td>
<td>2</td>
<td>33%</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>3</td>
<td>43%</td>
<td>2</td>
<td>28.5%</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>3</td>
<td>37.5%</td>
<td>3</td>
<td>37.5%</td>
</tr>
</tbody>
</table>

TABLE III
LIST OF SYMBOLS

xₖ, xₖ̂: Continuous and discrete state vectors, respectively, at time t or time step k
zₖ, ẑₖ: Continuous and discrete observation vectors, respectively, at time t or time step k
eₖ, êₖ: Continuous and discrete observation noise vectors, respectively, at time t or time step k
wₖ, ŵₖ: Continuous and discrete system process noise vectors, respectively, at time t or time step k
f(xₖ): Continuous-time system dynamics function
h(xₖ): Continuous-time observation function
U(xₖ): Negative gradient of the potential function
F_k: Jacobian matrix of the function f(x_k)
H_k: Jacobian matrix of the function h(x_k)
F̂_k: Discretized, linearized system transition matrix at time step k
Ĥ_k: Discretized, linearized observation matrix at time step k
R_k: Observation noise covariance matrix at time step k
W_k: System process noise covariance matrix at time step k
K_k: Kalman filter gain at time step k
δ̂_k: Error between predicted and true state vectors
δ̂_k: Redundant observation vector at time step k
δ_k: Noise vector corresponding to δ̂_k
R_k: Noise covariance matrix corresponding to δ̂_k
S_k: Cholesky Decomposition of A_k
F: Target probability distribution
G: c-contaminated probability distribution
H: Contamination probability distribution
c: Amount of contamination
T(G): Functional form of the estimator at G
r_k: ^th residual
h_k: Transpose of the i^th row of matrix H
Δ̂_k: Infinitesimal contamination at r
Σ̂_k: Predicted error covariance matrix at time step k
given data up to time step k – 1
Σ_k: Filter error covariance matrix at time step k given data up to time step k
IF: Influence function
x̂_k: Predicted state vector at time step k given data up to time step k – 1
x̂_k: Filtered state vector at time step k given data up to time step k
m_2: Total number of observations in x
n: Number of state variables in x
P_S: Projection Statistic for the i^th data point
q(·): Weight function used in a GM-estimator
ω_i: Weights derived from the projection statistics
y: Observation vector after whitening
A_k: Combined system and observation matrix corresponding to y
η: Noise vector corresponding to y
a_k: Transpose of the i^th row of the matrix A_k
J(·): Objective function for a GM-estimator
ρ(·): ρ-function used in GM-estimator
Ψ(·): Ψ-function of a GM-estimator
r_k: ^th scaled residuals
P: Target Gaussian probability distribution
Q: Diagonal weight matrix with diagonal elements, q(·)


**Mital A. Gandhi** (S05) received B.S. and M.S. in Electrical Engineering from the University of Illinois at Urbana-Champaign in 2002 and 2004, and a Ph.D. in Electrical Engineering from Virginia Tech in 2009. He is presently a Research Engineer at Lockheed Martin. His research interests include robust estimation, filtering, and acoustic signal processing.

**Lamine Mili** (S82, M88, SM93) received the B.S. degree from the Swiss Federal Institute of Technology, Lausanne, in 1976, and the Ph.D. degree from the University of Liege, Belgium, in 1987. He is presently a Professor of Electrical and Computer Engineering at Virginia Tech. His research interests include robust statistics, robust signal processing, radar systems, and power systems analysis and control.