Article

Identification of the Four-Bar Linkage Size in a Beam Pumping Unit Based on Cubature Kalman Filter

Jiaojian Yin 1,*, Dong Sun 2 and Hongzhang Ma 1

1 College of Science, China University of Petroleum (East China), Qingdao 266580, China
2 Technical Testing Center of Shengli Oilfield Branch of Sinopec, Dongying 257100, China
* Correspondence: yinjiaojian@upc.edu.cn

Abstract: While the size of the four-bar linkage is the basis of kinematic performance analysis in a beam pumping unit, there is still a lack of effective and direct measurement of it. Since the motor input power and the polished rod position are commonly used production data, a size identification algorithm of the four-bar linkage based on the motor input power and the polished rod position is proposed in this paper. Firstly, the kinematic model of a beam pumping unit, the speed model of a motor, the initial value, and the state space model are established. Secondly, a Cubature Kalman filter with nonlinear constraints is designed and the size identification algorithm is worked out. Lastly, the model and the size identification algorithm are validated based on the simulated and measured data, and the characteristics of identification with different measurement covariance are studied. The results demonstrate that both the model and the algorithm are feasible. The maximum relative error between the identified and the real size increases from 0.23% to 9.37% when the covariance increases from $10^{-6}$ to $10^{-1}$. With the measured covariance of the polished rod velocity, the maximum relative error is 7.09%. A comparison of several current identification methods demonstrates that the proposed algorithm is more accurate.

Keywords: four-bar linkage size; beam pumping units; Cubature Kalman filter; size identification algorithm

1. Introduction

As a major production equipment in the oilfield, the beam pumping unit enjoys huge installed capacity worldwide [1]. It possesses the advantages of simplicity, reliability, ease of operation and maintenance, and adaptability to harsh environments [2,3].

A sucker-rod pumping system is composed of the downhole pump, the sucker-rod string, the surface pumping unit, the gear reducer (gearbox), the V-belt drive, and the prime mover [4]. As a typical sucker-rod pumping system, the beam pumping unit has a distinctive surface pumping unit of a four-bar linkage, with the cranks, the linkage rod, walking beam, the standard bar, and the horsehead [5–7].

The four-bar linkage can convert the motor’s rotary motion, which is decelerated by the belt and gearbox drive into the oscillatory motion of the polished rod. It is one of the key mechanisms of the beam pumping unit. The size of the four-bar linkage is crucial to the contour size and the working performance of the beam pumping unit [8]. Only when its size is known can the position, velocity, and acceleration of the polished rod and the torque factors as functions of crank angle be calculated [7]. In addition, the crank radius also determines the balance of the beam pumping unit, which is closely related to the energy consumption or even the service life of the beam pumping unit [9]. Therefore, the size of the four-bar linkage is of vital importance for the design, behavior prediction, and state detection of a beam pumping unit. In recent years, some new technologies, such as the indirect measurement DC based on measured motor power [10], fault diagnosis based on motor power curves [11,12], prediction of submergence depth based on a hybrid model [5],
and variable speed drive technology of beam pumping units [13], require a more accurate size for the four-bar linkage.

However, it is difficult to measure the sizes of the four-bar linkage directly in actual production because of their large size and the continuous motion of the four-bar linkage. Sun et al. [14] proposed a kinematic analysis model of a beam pumping unit based on deep learning, which can give the size of the four-bar linkage. The method is built on the real-time video stream taken by the mobile robot, which is easily affected by the camera and the weather. At present, most of the size of the four-bar linkage is supplied by manufacturers. Obviously, it cannot reflect the actual size, because the size will change due to the equipment installation or a long-time operation. Since the polished rod position and the motor input power are the primary real-time data monitored continuously [10], it is necessary to measure the sizes of the four-bar linkage with a soft method. In this paper, an identification method will be proposed based on the measured polished rod position and the motor input power.

Parameter identification plays an important role in controller design, dynamic systems modeling, and signal processing [15–17]. More mainstream identification methods have been proposed for linear systems and nonlinear systems, such as the recursive least squares (RLS) methods [18,19], the stochastic gradient descent (SGD) methods [20,21], the artificial neural network (ANN) techniques [22,23], and the Kalman filter. For linear time invariant systems, the RLS method is widely used and easy implemented [24,25] but susceptible to noise [26]. The SGD method involves a small amount of calculation with a low accuracy [27]. The ANN technique benefits from higher accuracy as well as convenience in modeling but demands a training process [28].

Notably, the Kalman filter (KF) is typical, effective, and still widely used with the merits of real time, fast speed, and strong anti-interference ability [29]. The Kalman filter was proposed by Kalman in 1960 [30], and is only suitable for linear systems. With a view to apply the Kalman filter to nonlinear spacecraft navigation, the extended Kalman filter (EKF) was originally proposed by Stanley Schmidt in 1967 [31,32]. The basic idea of EKF is utilizing the value of first-order nonlinear Taylor expansion around the estimated status [33]. However, the EKF can lead to unstable results when the nonlinearity in the system is strong [34]. To overcome this theoretical deficiency, the unscented Kalman filter (UKF) was first proposed by Julier et al. in 1995 [35], with unscented transformation (UT) as its core [36]. Considering the deterioration and even divergence of the filter performance when the state dimension of the system is more than 3 in UKF, Arasaratnam and Haykin [37] proposed a Cubature Kalman filter (CKF) algorithm based on cubature transform, which is an unconstrained state Cubature Kalman filter. In practical applications, physical limitations always impose some constraints on the states or parameters of the system [38], so it is of great significance to study the constrained Cubature Kalman filter (CCKF). Simon [39] summarized the various ways to incorporate state constraints into the Kalman filter and its nonlinear modifications, such as model reduction, estimate projection, probability density function, etc., which provides a guideline to research the constrained Kalman filter. Zarei and Shokri [38] proposed a CCKF algorithm, which integrates linear constraints into the CKF by projecting unconstrained CKF estimation onto the boundary of the admissible region. However, it needed a Taylor series expansion of the constraint equation.

A brief comparison of different identification methods is tabulated in Table 1.

From the above survey, there is still a lack of an effective method to measure the four-bar linkage sizes of the beam pumping units while the identification method will be a feasible scheme. Due to the strong non-linearity of the four-bar linkage mechanism motion, the CKF is the best choice. With the consideration of the limitations in the sizes identification of four-bar linkage, a nonlinear constrained Cubature Kalman filter (NCCKF) without Taylor series expansion should be further studied. In this paper, a NCCKF is employed to identify the sizes of the four-bar linkage.
The contributions of this paper are as follows:

1. A speed model of a motor is established. Based on the nameplate parameters and the input power, the motor speed can be obtained.
2. An initial value model is established. Based on the measured polished rod position, the initial value and the lower boundary and upper boundary can be determined.
3. A NCCKF algorithm is proposed. Based on the maximized conditional probability density function, the predicted state is modified by solving the constrained optimization problem.
4. A size identification algorithm based on the measured motor input power and the polished rod position is proposed.

The rest of this paper is arranged as follows: Section 2 is the process of system model, Section 3 is the design of NCCKF, Section 4 is the validation and analysis of the size identification algorithm, Section 5 is the results discussion, and Section 6 is the conclusions of the paper.

2. System Model

2.1. Kinematic Model of A Pumping Unit

The geometric diagram of the four-bar linkage in a beam pumping unit is shown in Figure 1, where \( R \) is the radius of the crank, \( P \) is the length of the pitman, \( C \) is the length of the rear arm of walking beam, \( A \) is the length of the forearm of walking beam, and \( K \) is the length of the fixed bar.

![Figure 1. Geometric diagram of the four-bar linkage in a beam pumping unit.](image)

As is illustrated in Figure 1, when the crank rotates clockwise and the polished rod starts from the bottom of the stroke, it has

\[
\left\{
\begin{array}{l}
\theta_1 = \arccos\left(\frac{(P+R)^2 + K^2 - C^2}{2(P+R)K}\right) \\
\theta_2(t) = 2\pi + \theta_1 - \theta(t)
\end{array}
\right.
\]

(1)
According to Figure 1, it has the following equation [7]:

\[
Re^{i\theta_2(t)} + Pe^{i\theta_3(t)} = K + Ce^{i\theta_4(t)}
\]  

(2)

When the sizes of \( R, C, P, \) and \( K \) are known, the angles of \( \theta_3(t) \) and \( \theta_4(t) \) can be solved according to the following equation:

\[
\begin{cases}
R \cos \theta_2(t) + P \cos \theta_3(t) - K - C \cos \theta_4(t) = 0 \\
R \sin \theta_2(t) + P \sin \theta_3(t) - C \sin \theta_4(t) = 0
\end{cases}
\]

(3)

Thus, the polished rod position is obtained by

\[
\begin{cases}
P_K(t) = A[\theta_4(t) - \theta_{4\text{min}}] \\
\theta_{4\text{min}} = \pi - \arccos \frac{C^2 + K^2 - (P + P_r)^2}{2CK}
\end{cases}
\]

(4)

The polished rod velocity and acceleration are given by

\[
\begin{cases}
v_p(t) = A\dot{\theta}_4(t) \\
a_p(t) = A\ddot{\theta}_4(t)
\end{cases}
\]

(5)

where [7]

\[
\begin{cases}
\dot{\theta}_2(t) = -\dot{\theta}(t) \\
\dot{\theta}_3(t) = \frac{R\dot{\theta}_3(t) \sin \theta_3(t) - \dot{\theta}_2(t)}{P \sin \theta_3(t) - \dot{\theta}_4(t)} \\
\dot{\theta}_4(t) = \frac{R\dot{\theta}_3(t) \sin \theta_3(t) - \dot{\theta}_2(t)}{C \sin \theta_3(t) - \dot{\theta}_4(t)}
\end{cases}
\]

(6)

\[
\begin{cases}
\ddot{\theta}_2(t) = -\ddot{\theta}(t) \\
\ddot{\theta}_3(t) = \frac{\ddot{\theta}_2(t) - \dot{\theta}_2(t) \cot \theta_3(t) - \dot{\theta}_4(t) + \ddot{\theta}_3(t) \cot \theta_2(t) (\dot{\theta}_3(t) - \dot{\theta}_4(t))}{\ddot{\theta}_2(t) - \dot{\theta}_2(t) \cot \theta_3(t) - \dot{\theta}_4(t) + \ddot{\theta}_3(t) \cot \theta_2(t) (\dot{\theta}_3(t) - \dot{\theta}_4(t))}
\end{cases}
\]

(7)

Therefore, the established kinematic model of a pumping unit is as follows:

\[
\begin{cases}
P_K(t) = f_p[R, P, C, K, A, \theta(t)] \\
v_K(t) = f_v[R, P, C, K, A, \theta(t), \dot{\theta}(t)] \\
a_K(t) = f_a[R, P, C, K, A, \theta(t), \dot{\theta}(t), \ddot{\theta}(t)]
\end{cases}
\]

(8)

If the crank rotates anticlockwise, it has

\[
\begin{cases}
\theta_2(t) = \theta_1 + \theta(t) \\
\dot{\theta}_2(t) = \dot{\theta}(t) \\
\ddot{\theta}_2(t) = \ddot{\theta}(t)
\end{cases}
\]

(9)

Thus, if the kinematic data of the crank, i.e., angle, angular velocity, and angular acceleration, are known, the sizes of \( R, C, P, K, \) and \( A \) can be identified according to the measured polished rod kinematic data—that is, the polished rod position, velocity, and acceleration. In actual production, the measured data are the polished rod position, and the polished rod velocity and acceleration can be obtained by differentiating them.

### 2.2. Speed Model of A Motor

Most of the prime movers of the beam pumping units are three-phase asynchronous motors (referred to as motor for convenience). When the belt drive does not slip, the crank angle will be proportional to the motor output shaft angle. Thus, if the motor speed can be determined, the crank angle, velocity, and acceleration can be calculated. Since it is easy to obtain the input power of a motor in real-time [10], it is necessary to establish a motor speed model based on the input power.
As is shown in Figure 2 [40,41], most of the equivalent circuit of a motor is T-circuit. When the values of the circuit parameters $R_1$, $R_2$, $X_1$, $X_2$, and $X_m$, are given, the motor torque and the input power can be calculated as follows without consideration of the mechanical loss:

$$T_2(t) = \frac{3}{\omega_s} |I_2(t)|^2 \frac{R_2}{s(t)}$$  \hspace{1cm} (10)

$$P_{in}(t) = 3 \text{real} [I_1(t)] U_{ph}$$  \hspace{1cm} (11)

where

$$Z_{in}(t) = R_1 + jX_1 + \left[ \frac{jX_m \frac{R_2}{s(t)} + jX_2}{jX_m + j(X_m + X_2)} \right]$$  \hspace{1cm} (12)

$$I_1(t) = \frac{U_{ph}}{Z_{in}(t)}$$

$$I_2(t) = I_1(t) \frac{X_m}{jX_m + j(X_m + X_2)}$$

$$\omega_s = \frac{2\pi n_s}{60}$$  \hspace{1cm} (13)

and $n_s$ is the synchronous speed of the motor.

![Figure 2. T-type reduced to one-phase equivalent circuit of a motor.](image)

As is shown in Equations (10) and (11), and Figure 2, the motor torque and the input power are rather complicated with the T-circuit because of its two serial–parallel branches. Thevenin approach can be used to establish the speed model and it is shown in Figure 3.

![Figure 3. Thevenin’s one-phase equivalent circuit for a motor.](image)

Circuit parameters $U_{Th}$, $R_{Th}$, and $X_{Th}$ are determined with the Kirchhoff circuit rules as follows [40]:

$$\begin{cases}
U_{Th} = \epsilon U_{ph} \\
R_{Th} = \epsilon^2 R_1 \\
X_{Th} = X_1 \\
\epsilon = \frac{X_m}{X_1 + X_m}
\end{cases}$$  \hspace{1cm} (14)

According to Figure 3, the rotor current is given by

$$I_{Th}(t) = \frac{U_{Th}}{R_{Th} + \frac{R_2}{s(t)} + j(X_{Th} + X_2)}$$  \hspace{1cm} (15)
Thus, the power dissipated across the stator resistance $R_{Th}$ is given by

$$P_{Cu}(t) = 3|I_{Th}(t)|^2 R_{Th}$$  \hspace{1cm} (16)$$

The total power transferred across the air gap from the stator is given by

$$P_M(t) = 3|I_{Th}(t)|^2 \frac{R_2}{s(t)}$$  \hspace{1cm} (17)$$

Without consideration of the mechanical loss, the total input power is given by

$$P_{in}(t) = P_{Cu}(t) + P_M(t)$$  \hspace{1cm} (18)$$

According to Equation (15), Equations (16) and (17) can be rewritten as:

$$P_{Cu}(t) = \frac{3U_{Th}^2 R_{Th}}{\left[ R_{Th} + \frac{R_2}{s(t)} \right]^2 + (X_{Th} + X_2)^2}$$  \hspace{1cm} (19)$$

$$P_M(t) = \frac{3U_{Th}^2 R_2}{\left[ R_{Th} + \frac{R_2}{s(t)} \right]^2 + (X_{Th} + X_2)^2}$$  \hspace{1cm} (20)$$

Thus, it has the maximum power $P_{Mm}$ and the maximum slide $s_m$ in the motor state [42] according to Equation (20):

$$\begin{align*}
P_{Mm} &= \frac{3\frac{U_{Th}^2}{R_2}}{2 R_{Th} + \frac{R_2}{s_m}} \\
s_m &= \sqrt{R_{Th} + (X_{Th} + X_2)^2}
\end{align*}$$  \hspace{1cm} (21)$$

According to Equation (21), Equations (19) and (20) can be written as:

$$P_{Cu}(t) = \frac{2P_{Mm}(\sigma s_m + 1)\sigma s(t)}{\sigma s(t) + \sigma s_m + 2\sigma s_m}$$  \hspace{1cm} (22)$$

$$P_M(t) = \frac{2P_{Mm}(\sigma s_m + 1)}{\sigma s_m + \sigma s(t) + 2\sigma s_m}$$  \hspace{1cm} (23)$$

where

$$\sigma = \frac{R_{Th}}{R_2}$$  \hspace{1cm} (24)$$

Thus, the total input power is given by

$$P_{in}(t) = \frac{2P_{Mm}(\sigma s_m + 1)[\sigma s(t) + 1]}{\sigma s(t) + \sigma s_m + 2\sigma s_m}$$  \hspace{1cm} (25)$$

Resolving Equation (25) yields the slide

$$s(t) = \frac{\left[ \sigma P_{in}(t)s_m^2 - \frac{1}{2} \right] - \sqrt{\left[ \sigma P_{in}(t)s_m^2 - \frac{1}{2} \right]^2 - P_{in}(t)s_m^2 [P_{in}(t) - \sigma]}}{P_{in}(t) - \sigma}$$  \hspace{1cm} (26)$$

where

$$P_{in} = \frac{P_n}{2P_{Mms_m}(\sigma s_m + 1)}$$  \hspace{1cm} (27)$$
Therefore, it has the motor speed:

\[
n(t) = n_s + n_s \frac{\left[ \sigma P_{\text{in}}(t) \dot{s}_m^2 - \frac{1}{2} \right] + \sqrt{\left[ \sigma P_{\text{in}}(t) \dot{s}_m^2 - \frac{1}{2} \right]^2 - P_{\text{in}}(t) s_m^2 P_{\text{in}}(t) - \sigma}}{P_{\text{in}}(t) - \sigma}
\]

Thus, the motor speed can be calculated based on the input power when the parameters of \( P_{\text{Mm}}, S_m, \) and \( \sigma \) are given. There are two methods to determine these parameters. One is to calculate it based on Equations (21) and (24) according to the circuit parameters, the other is according to the nameplate parameters. The nameplate parameters include the rated power \( P_H \), rated slide \( S_H \), rated efficiency \( \eta_H \), and maximum torque ratio \( \lambda_k \). It has

\[
\begin{align*}
P_{\text{Mm}} &= \lambda_k P_H \\
S_m &= S_H \left( \lambda_k + \sqrt{\lambda_k^2 - 1} \right)
\end{align*}
\]

To determine \( \sigma \), the rated efficiency \( \eta_H \) should be used. In view of the output motor power, it has:

\[
P_{\text{Mot}} = (1 - s) P_M
\]

For \( \eta = \frac{P_{\text{Mot}}}{P_m} \), it has the normal efficiency according to Equations (25) and (30)

\[
\eta_H = \frac{1 - s_H}{s_H + 1}
\]

Therefore, it has:

\[
\sigma = \frac{1 - s_H - \eta_H}{\eta_H S_H}
\]

Since the crank angle is proportional to the motor output shaft angle, it has the angular velocity as follows:

\[
\begin{align*}
\dot{\theta}(t) &= i_{MB} \dot{\theta} \frac{2\pi n(t)}{60} \\
\dot{\theta}_0 &= \frac{2\pi n_p}{60}
\end{align*}
\]

where \( i_{MB} \) is a transmission ratio and \( n_p \) is the pumping speed.

Thus, the crank angle can be obtained by integrating the angular velocity, and the acceleration can be obtained by differentiating the angular velocity.

### 2.3. Initial Value and State Space Model

The initial value of the parameter is very important for the parameter identification. According to the movement of the crank slider mechanism, the polished rod position can be approximated as follows [43]:

\[
P_{Ra}(t) = \lambda_1 \left[ 1 - \cos \theta(t) + \frac{\lambda_1}{2} \sin^2 \theta(t) \right]
\]

where

\[
\begin{align*}
\lambda_1 &= AR/C \\
\lambda_2 &= R/P
\end{align*}
\]

The values of the \( \lambda_1 \) and \( \lambda_2 \) can be determined according to the optimization problems:

\[
\min f(\lambda_1, \lambda_2) = \sqrt{\sum_{i=1}^{N_i} \left[ P_R(\theta_i) - \tilde{P}_{Ra}(\theta_i) \right]^2}
\]

where \( \theta(N_i) = 0.8 \theta_m \), \( \theta_m \) is the crank angle of the maximum polished rod position.
In the design of the beam pumping unit, the maximum crank radius \( R_0 \), the length of the pitman, and the rear arm of the walking beam have the relative size restrictions, that is [44]:

\[
\begin{align*}
0.35 & \leq R_0 / P \leq 0.4 \\
0.45 & \leq R_0 / C \leq 0.6 
\end{align*}
\] (37)

If the intermediate value in Equation (37) is taken, it has an initial value of \( R_0 / C \) according to Equation (35), that is, \( R_0 / C = 1.4 \lambda_2 \).

Thus, the initial value of the identification can be calculated according to the symmetrical cycle working mode of a beam pumping unit [44].

\[
\begin{align*}
A_0 & = P R_m^2 \arcsin \left( \frac{R_0}{C} \right) = P R_m^2 \arcsin (1.4 \lambda_2) \\
C_0 & = A_0 / (A / C) = A_0 / 1.5 \\
R_0 & = C_0 (R_0 / C) = 1.4 \lambda_2 C_0 \\
P_0 & = R_0 / (R / P) = R_0 / \lambda_2 \\
K_0 & = \sqrt{P_0^2 + C_0^2 - R_0^2} 
\end{align*}
\] (38)

where \( P_R m \) is the maximum value of the polished rod position.

In the design of the beam pumping unit, the relative size restrictions among each rod length of the four-bar linkage are as follows [45]:

\[
\begin{align*}
0.7 & \leq P / K \leq 0.85 \\
0.4 & \leq C / K \leq 0.7 \\
1 & \leq A / C \leq 2.0 
\end{align*}
\] (39)

According to the resolution of optimization problems of Equation (36) and the restriction of Equation (39), its restriction is as follows:

\[
\begin{align*}
0.5 \lambda_1 & \leq R \leq \lambda_1 \\
0.23 (\lambda_1 / \lambda_2) & \leq C \leq (\lambda_1 / \lambda_2) \\
0.5 (\lambda_1 / \lambda_2) & \leq P \leq (\lambda_1 / \lambda_2) \\
0.58 (\lambda_1 / \lambda_2) & \leq K \leq 1.43 (\lambda_1 / \lambda_2) \\
0.23 (\lambda_1 / \lambda_2) & \leq A \leq 2 (\lambda_1 / \lambda_2) 
\end{align*}
\] (40)

To avoid complex numbers in angle calculations, it has the nonlinear constraint:

\[
\begin{align*}
\left| \frac{C^2 + K^2 - (P - R)^2}{2CK} \right| & \leq 1 \\
R + K & \leq C + P 
\end{align*}
\] (41)

In this paper, the identified parameter is the size of the four-bar linkage, and the measured data are the polished rod position, velocity, and acceleration. Therefore, it has the state and measured vector of

\[
\begin{align*}
\{x(t) = [R \ C \ P \ K \ A]^T \\
z(t) = [P_p(t) \ v_R(t) \ a_R(t) \ 0 \ 0]^T 
\end{align*}
\] (42)

The state equation and the measurement equation are as follows:

\[
\begin{align*}
\{x(t) = Fx(t) + \omega(t) \\
z(t) = Fh[x(t)] + \nu(t) 
\end{align*}
\] (43)

where

\[
\begin{align*}
F &= I_5 \\
h[x(t)] &= \begin{bmatrix} f_{p}(x(t), \theta(t)) & f_{v}(x(t), \theta(t), \dot{\theta}(t)) & f_{a}(x(t), \theta(t), \dot{\theta}(t), \ddot{\theta}(t)) & 0 & 0 \end{bmatrix}^T 
\end{align*}
\] (44)
where $w(t)$ and $v(t)$ are the independent process and the observation Gaussian noise with zeros means and covariances $Q_c$ and $R_c$, respectively. $I_5$ is 5-dimensional identity matrix.

According to Equation (38), the initial value is

$$x_0 = \begin{bmatrix} R_0 & C_0 & P_0 & K_0 & A_0 \end{bmatrix}^T$$

(45)

And the nonlinear constraint is

$$\begin{cases} x_L \leq x \leq x_U \\ g(x) \leq 0 \end{cases}$$

(46)

where

$$\begin{align*}
  x_L &= \frac{\lambda_1}{\lambda_2} \begin{bmatrix} 0.50 \lambda_2 & 0.23 & 0.50 & 0.58 & 0.23 \end{bmatrix}^T \\
  x_U &= \frac{\lambda_1}{\lambda_2} \begin{bmatrix} \lambda_2 & 1 & 1 & 1.43 & 2 \end{bmatrix}^T \\
  g(1) &= \frac{x(2)^2+x(4)^2-(x(3)-x(1))^2}{2x(2)x(4)} - 1 \\
  g(2) &= x(1) + x(4) - x(2) - x(3)
\end{align*}$$

(47)

After the state equation is discretized, it has:

$$\begin{cases} x_{k+1} = Fx_k + w_k \\ z_{k+1} = Fh_{k+1}(x_{k+1}) + v_{k+1} \end{cases}$$

(48)

3. Design of Cubature Kalman Filter with Nonlinear Constraints

This section aims to find a Cubature Kalman filter algorithm with nonlinear constraint to identify the size of the four-bar linkage.

3.1. Cubature Kalman Filter without Constraint

The identification in this paper is a nonlinear problem and it has five state variables. Therefore, the Cubature Kalman filter is selected.

The Cubature Kalman filter is a Bayesian filter theory in the Gaussian domain and a set of volume points is employed to approximate the state mean and covariance of non-linear systems based on the third-order spherical radial cubature criterion [37].

The CKF algorithm is summarized below [37,38]:

Step 1. Initialize with

$$\hat{x}_0 = E(x_0)$$
$$P_0 = E\left[(x_0-\hat{x}_0)(x_0-\hat{x}_0)^T\right]$$

Step 2. Decompose covariance by Cholesky

$$S_{k-1|k-1} = \text{chol}(P_{k-1|k-1})$$

Step 3. Evaluate the cubature points, $i = 1, 2, ..., m$ (where $m = 2n$):

$$X_{i,k-1|k-1} = S_{k-1|k-1}\xi_i + \hat{x}_{k-1|k-1}$$

where $\xi_i$ is the i-th element of the following set [46]:

$$\xi_i = \sqrt{m/2}\delta_i$$

where $[\delta] = [I_n, -I_n]$, and $I_n$ is n-dimensional identity matrix.

Step 4. Evaluate the propagated cubature points, $i = 1, 2, ..., m$ (where $m = 2n$):

$$X_{i,k|k-1} = FX_{i,k-1|k-1}$$
Step 5. Estimate the predicted state and predicted error covariance:
\[
\hat{x}_{k|k-1} = \frac{1}{m} \sum_{i=1}^{m} X_{i,k|k-1}^T
\]
\[
P_{k|k-1} = \frac{1}{m} \sum_{i=1}^{m} X_{i,k|k-1}^T X_{i,k|k-1} - \hat{x}_{k|k-1} \hat{x}_{k|k-1}^T + Q_{r,k-1}
\]

Step 6. Decompose covariance by Cholesky:
\[
S_{k|k-1} = chol\left(P_{k|k-1}\right)
\]

Step 7. Redraw cubature points:
\[
X_{i,k|k-1} = S_{k|k-1} z_i + \hat{x}_{k|k-1}
\]

Step 8. Estimate the predicted measurement and associated covariance:
\[
Z_{i,k|k-1} = F h \left(X_{i,k|k-1}\right)
\]
\[
\tilde{z}_{k|k-1} = \frac{1}{m} \sum_{i=1}^{m} Z_{i,k|k-1}
\]
\[
P_{zz,k|k-1} = \frac{1}{m} \sum_{i=1}^{m} Z_{i,k|k-1} Z_{i,k|k-1}^T - \tilde{z}_{k|k-1} \tilde{z}_{k|k-1}^T + R_{z,k}
\]
\[
P_{xz,k|k-1} = \frac{1}{m} \sum_{i=1}^{m} X_{i,k|k-1} Z_{i,k|k-1}^T - \tilde{x}_{k|k-1} \tilde{z}_{k|k-1}^T
\]

Step 9. Estimate the cubature Kalman gain:
\[
W_k = P_{xz,k|k-1} P_{zz,k|k-1}^{-1}
\]

Step 10. Update the state estimation and associated covariance:
\[
\hat{x}_k = \hat{x}_{k|k-1} + W_k \left(z_k - \tilde{z}_{k|k-1}\right)
\]
\[
P_k = P_{k|k-1} - W_k P_{zz,k|k-1} W_k^T
\]

Step 11. Go to step 2 for the next sample.

### 3.2. Cubature Kalman Filter with Nonlinear Constraint

The essence of the Kalman filter is that its estimate is that value of \(x_k\) that maximizes the conditional probability density function \(P(x_k|z_k)\) [47]:

\[
P(x_k|z_k) = \frac{\exp\left[-\left(x_k - \hat{x}_{k|k}\right) P_{k|k}^{-1} \left(x_k - \hat{x}_{k|k}\right)^T / 2\right]}{(2\pi)^n/2 |P_{k|k}|^{1/2}}
\]  \hspace{1cm} (49)

Therefore, if the Kalman filter is constrained, an estimate \(\tilde{x}_k\) should be found to maximize the conditional probability \(P(\tilde{x}_k|z_k)\) with \(\tilde{x}_k\) satisfying the constraint [47], so it has the constrained optimization problems in this paper:

\[
\begin{align*}
\min \left(\tilde{x}_k - \hat{x}_{k|k}\right) P_{k|k}^{-1} \left(\tilde{x}_k - \hat{x}_{k|k}\right)^T \\
x_k \leq \tilde{x}_k \leq x_H \\
g(\tilde{x}_k) \leq 0
\end{align*}
\]  \hspace{1cm} (50)

Notably, when \(P_{k|k}^{-1}\) is replaced by any symmetric positive definite weighting matrix \(W\), the maximum probability method is the projection method. With consideration of the characteristics of cubature transform, the projection method of cubature points in the
Cubature Kalman filter [37] can be modified, and the checking and modifying between step 6 and step 7 of the ordinary CKF algorithm should be added. It is NCCKF step 1. Check the state estimation

if \( \hat{x}_{k|k-1} \) is out of constraint,
resolve constrained optimization problem:

\[
\begin{align*}
\min \left( \hat{x}_k - \hat{x}_{k|k-1} \right) & \cdot P^{-1}_{k|k-1} \left( \hat{x}_k - \hat{x}_{k|k-1} \right)^T \\
x_L & \leq \hat{x}_k \leq x_H \\
g_1(\hat{x}_k) & \leq 0 \\
g_2(\hat{x}_k) & \leq 0
\end{align*}
\]

(51)

And update the state, that is \( \hat{x}_{k|k-1} = \hat{x}_k \).
Otherwise, turn to the step 7 of the ordinary CKF algorithm.

The optimization problem of Equation (51) can be easily resolved by the active-set method.

The size identification algorithm is summarized in Algorithm 1.

**Algorithm 1.** Size identification of a four-bar linkage.

1. Calculate \( \theta(t) \) according to (28) and (33), then calculate \( \hat{\theta}(t) \) and \( \hat{\theta}(t) \)
2. Obtain \( \theta_{m} \), calculate \( \lambda_1 \) and \( \lambda_2 \) according to (36) and then obtain \( x_0, x_L, \) and \( x_H \)
3. Initialize \( w, v, P_0, \hat{x}_0 = x_0 \)
4. for \( k = 2:N \)
5. Decompose covariance \( P_{k-1|k-1} \)
6. Evaluate the cubature points and obtain \( X_{i,k-1|k-1} \)
7. Evaluate the propagated cubature points and obtain \( X'_{i,k|k-1} \)
8. Estimate the predicted state and predicted error covariance, then obtain \( \hat{x}_{k|k-1}, P_{k|k-1} \)
9. Decompose covariance \( P_{k|k-1} \) and obtain \( S_{k|k-1} \)
10. Check the state estimation

   if \( \hat{x}_{k|k-1} \) is out of constraint

   Resolve constrained optimization problem (51) and update the state \( \hat{x}_{k|k-1} = \hat{x}_k \)
end
11. Redraw cubature points and obtain \( X'_{i,k|k-1} \)
12. Estimate the predicted measurement and associated covariance, then obtain

\[
Z_{i,k|k-1} = Z_{i,k|k-1}^* P_{zz,k|k-1} = P_{zz,k|k-1}^* P_{zz,k|k-1}
\]
13. Estimate the cubature Kalman gain and obtain \( W_k \)
14. Update the state estimation and associated covariance, then obtain \( \hat{x}_{k|k}, P_{k|k} \)
15. End
16. Output and plot \( z, \hat{z}, \hat{x} \)

4. Validation and Analysis

In this section, the size identification algorithm will be validated based on the simulated and measured data. The first is the validation of the speed model based on the input power simulated by the T-type circuit motor model, the nameplate parameters and circuit parameters are given. The second is the validation of the size identification algorithm based on simulated the polished rod position, velocity, and acceleration. The third is the validation of the NCCKF algorithm based on the simulated polished rod position, velocity, and acceleration. The last is the validation of the size identification and result analysis based on the measured motor input power and the polished rod position of a beam pumping unit. MATLAB R2021a code is applied in this paper.

4.1. Validation of Speed Model

The nameplate parameters of the motor are shown in Table 2 and its T-type equivalent circuit parameters are \( R_1 = 0.3958 \, \Omega \), \( X_1 = 0.4002 \, \Omega \), \( R_2 = 0.8837 \, \Omega \), \( X_2 = 2.1085 \, \Omega \), and
\( X_m = 13.0658 \, \Omega \), respectively. Thus, the parameters of the Thevenin equivalent circuit can be calculated according to Equation (14).

### Table 2. Nameplate parameters of the motor.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>poles</td>
<td>6</td>
<td>( P_{\text{nom}}, \text{kW} )</td>
<td>13.3</td>
</tr>
<tr>
<td>( n_s, \text{rpm} )</td>
<td>1000</td>
<td>( \eta_{\text{nom}} )</td>
<td>0.743</td>
</tr>
<tr>
<td>( n_{\text{nom}}, \text{rpm} )</td>
<td>830</td>
<td>( \cos \varphi_{\text{nom}} )</td>
<td>0.84</td>
</tr>
<tr>
<td>( (U_{ph})_{\text{nom}}, \text{V} )</td>
<td>380</td>
<td>( \lambda_k )</td>
<td>3.34</td>
</tr>
<tr>
<td>( I_{\text{nom}}, \text{A} )</td>
<td>32</td>
<td>connection</td>
<td>star (Y)</td>
</tr>
</tbody>
</table>

In view of the fluctuating load of the beam pumping unit, the output torque of the motor is set as:

\[
T(t) = 46.42 - 22.40 \cos(2\dot{\theta}_0 t) - 18.40 \cos(3\dot{\theta}_0 t) + 10.40 \sin(3\dot{\theta}_0 t) \quad (52)
\]

The pumping speed of the beam pumping unit is 1.7 times/min.

According to Equations (10) and (11), the input power of the T-type equivalent circuit motor model can be calculated and shown in Figure 4. Thus, the speed can be calculated based on the circuit parameters and Equations (21), (24) and (28). For comparison purposes, the above speed is defined as circuit speed. The other speed can be calculated according to Equations (28), (29) and (32) based on the nameplate parameters, which are defined as the nameplate speed for comparison purposes. The results are all shown in Figure 4. As is illustrated in Figure 4, the nameplate speed is consistent with the circuit speed and the relative error is 0.12%, which indicates that the speed model is effective. Here, the relative error represents the ratio of the maximum error between the nameplate speed and the circuit speed to the maximum value of the circuit speed.

![Figure 4. Simulated input power and speed curve.](image)

### 4.2. Validation of Size Identification Algorithm Based on the Simulated Data

The type of the beam pumping unit is CYJY10-4.2-53HB. The real values of the size are shown in Table 3.
Table 3. Real and identified size identification.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Real Value</th>
<th>Identified Value</th>
<th>Relative Error %</th>
</tr>
</thead>
<tbody>
<tr>
<td>R, m</td>
<td>1.1000</td>
<td>1.1479</td>
<td>−4.36</td>
</tr>
<tr>
<td>C, m</td>
<td>2.5000</td>
<td>2.6220</td>
<td>−4.88</td>
</tr>
<tr>
<td>P, m</td>
<td>3.6000</td>
<td>3.7703</td>
<td>−4.73</td>
</tr>
<tr>
<td>K, m</td>
<td>4.7424</td>
<td>4.9765</td>
<td>−4.94</td>
</tr>
<tr>
<td>A, m</td>
<td>4.4000</td>
<td>4.4208</td>
<td>−0.48</td>
</tr>
</tbody>
</table>

The crank velocity \( \dot{\theta}(t) \) can be calculated according to the nameplate motor speed shown in Figure 4 when \( i_{MB} \) is 0.01 according to Equation (33). Thus, the crank angle \( \theta(t) \) and acceleration \( \ddot{\theta}(t) \) can be obtained by integrating and differentiating the crank velocity. They are all shown in Figure 5.

![Figure 5. Calculated crank angle, velocity, and acceleration.](image)

Based on the crank angle, velocity, and acceleration, the polished rod position \( P_R(t) \), velocity \( v_R(t) \) and acceleration \( a_R(t) \) can be simulated according to Equation (8) with the real values of the size listed in Table 3. The simulated results are shown in Figure 6. The Equation (3) can be solved by the `fsolve` function of MATLAB. Now, it has the measurement series for the size identification, that is \( P_R(t), v_R(t), \) and \( a_R(t) \).

![Figure 6. Simulated polished rod position, velocity, and acceleration.](image)
According to the polished rod position shown in Figure 6, it has the maximum value at position of 3.3796 rad and the series length N is 118. The solution of Equation (36) with the \textit{fminunc} function of MATLAB obtains the value of \(\lambda_1\) and \(\lambda_2\). The simulated polished rod position, the redrawn position based on Equation (34) with the real values, and that based on Equation (34) with the solved values of \(\lambda_1\) and \(\lambda_2\) are shown in Figure 7. As is illustrated in Figure 7, the real value position is consistent with the solved value position and the relative error is 1.12%, which shows that the method to obtain the values of \(\lambda_1\) and \(\lambda_2\) by solving Equation (36) is feasible.

![Figure 7: Simulated, real, and solved polished rod position.](image)

So, the initial value, lower boundary, and up boundary can be calculated according to Equations (38) and (40) as follows

\[
\begin{align*}
x_0 &= \begin{bmatrix} 1.3554 & 2.8221 & 3.9509 & 4.6623 & 4.2331 \end{bmatrix}^T \\
x_L &= \begin{bmatrix} 0.9708 & 1.3316 & 2.8299 & 3.3291 & 1.3318 \end{bmatrix}^T \\
x_U &= \begin{bmatrix} 1.9417 & 5.6598 & 5.6598 & 8.0856 & 11.3197 \end{bmatrix}^T
\end{align*}
\]

When \(Q_c = 10^{-7}I_5\), \(R_c = 10^{-3}I_5\) and \(P_0 = 0.15I_5\), the size of the four-bar linkage can be identified according to the Algorithm 1. Checking the calculation process shows that the algorithm can run without constraints under this initial value. So, the algorithm is based on the ordinary CKF algorithm. The outputs of \(z\) and \(\dot{z}\), that is the simulated and predicted measurement data, are plotted in Figure 8, which shows that simulated and predicted polished rod position and velocity have good consistency. Predicted polished rod acceleration has a disturbance at the beginning, but it is updated to the simulation acceleration soon. The output of \(\ddot{z}\), that is identified size results of four-bar linkage, is plotted in Figure 9. For comparison, the real size values are also plotted in Figure 9. Figure 9 shows that the identified sizes can be from their initial values to the corresponding real values within the first 10 s. The last values of \(\ddot{z}\) are shown in Table 3. For comparison, the relative error between the identified and real values is also shown in Table 3. As listed in Table 3, the relative error of \(K\) is the maximum of 4.94% and the relative error of \(A\) is the minimum with 0.48%. Interestingly, the relative errors of \(R, C, P,\) and \(K\) are basically equal. The reason is that the polished rod position is only sensitive to the ratios between the sizes shown in Equations (3) and (4). The above results show that the algorithm proposed in this paper is feasible.
Figure 8. Simulated and predicted measurement data.

Figure 9. Identified size results.
To illustrate the affection of the measurement covariance $R_c$, the size of the four-bar linkage is identified according to the Algorithm 1 when $R_c$ increases from $10^{-6}$ to $10^{-1}$, that is, $R_{c1} = 10^{-6}$, $R_{c2} = 10^{-3}$, $R_{c3} = 10^{-2}$, $R_{c4} = 10^{-3}$, $R_{c5} = 10^{-2}$, and $R_{c6} = 10^{-1}$.

The relative error between the identified and the real values is shown in Table 4. As is listed in Table 4, with the increase in measurement covariance $R_c$, the relative error rises, which indicates that the accuracy of identification is reduced. Under the same covariance, the relative error of $A$ is the minimum while that of $R$, $C$, $P$, and $K$ is almost the same when the covariance is less than $10^{-1}$.

Table 4. Relative error with different measurement covariances.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$R_{c1}$ Error, %</th>
<th>$R_{c2}$ Error, %</th>
<th>$R_{c3}$ Error, %</th>
<th>$R_{c4}$ Error, %</th>
<th>$R_{c5}$ Error, %</th>
<th>$R_{c6}$ Error, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$, m</td>
<td>0.26</td>
<td>−1.17</td>
<td>−5.49</td>
<td>−4.36</td>
<td>−5.47</td>
<td>−9.30</td>
</tr>
<tr>
<td>$C$, m</td>
<td>0.29</td>
<td>−1.12</td>
<td>−6.23</td>
<td>−4.88</td>
<td>−4.81</td>
<td>−6.71</td>
</tr>
<tr>
<td>$P$, m</td>
<td>0.27</td>
<td>−1.32</td>
<td>−6.18</td>
<td>−4.73</td>
<td>−4.64</td>
<td>−3.62</td>
</tr>
<tr>
<td>$K$, m</td>
<td>0.28</td>
<td>−1.24</td>
<td>−6.33</td>
<td>−4.94</td>
<td>−4.81</td>
<td>−4.49</td>
</tr>
<tr>
<td>$A$, m</td>
<td>0.03</td>
<td>0.03</td>
<td>−0.73</td>
<td>−0.47</td>
<td>0.84</td>
<td>3.31</td>
</tr>
</tbody>
</table>

4.3. Validation of NCCKF Algorithm

If the initial value is inappropriate, the algorithm cannot run due to the arccosine operation in the model, and therefore the NCCKF algorithm is necessary. For the sake of validation of the NCCKF algorithm, $K_0 = 2P_0$, $Q_c = 10^{-7}I_5$, and $Q_0 = 0.15I_5$. Obviously, the ordinary CKF algorithm cannot run because the initial value is out of the constrained—that is, $g(1) > 0$ and $g(2) > 0$. The constrained optimization problem (51) is solved by fmincon function of MATLAB. When the covariances $R_c$ increases from $10^{-6}$ to $10^{-1}$, the identified size is shown in Table 5.

Table 5. Relative error with different measurement covariances.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$R_{c1}$ Error, %</th>
<th>$R_{c2}$ Error, %</th>
<th>$R_{c3}$ Error, %</th>
<th>$R_{c4}$ Error, %</th>
<th>$R_{c5}$ Error, %</th>
<th>$R_{c6}$ Error, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$, m</td>
<td>−36.90</td>
<td>−33.12</td>
<td>−37.72</td>
<td>−38.21</td>
<td>−36.64</td>
<td>4.55</td>
</tr>
<tr>
<td>$C$, m</td>
<td>−36.32</td>
<td>−33.31</td>
<td>−38.42</td>
<td>−40.48</td>
<td>−40.95</td>
<td>2.60</td>
</tr>
<tr>
<td>$P$, m</td>
<td>−35.87</td>
<td>−33.27</td>
<td>−38.42</td>
<td>−40.47</td>
<td>−40.55</td>
<td>−24.11</td>
</tr>
<tr>
<td>$K$, m</td>
<td>−35.87</td>
<td>−33.31</td>
<td>−38.54</td>
<td>−40.85</td>
<td>−41.54</td>
<td>−21.20</td>
</tr>
<tr>
<td>$A$, m</td>
<td>0.55</td>
<td>−0.15</td>
<td>−0.52</td>
<td>−1.72</td>
<td>−3.15</td>
<td>6.42</td>
</tr>
</tbody>
</table>

As is shown in Table 5, the result follows the same rule as the ordinary CKF with the increase in covariances while the relative error of $R$, $C$, $P$, and $K$ is almost 40%, which indicates that the algorithm is feasible but only converges to a local optimum value instead of the real value. If the ratio of $A$ and $C$ (or anyone among $R$, $P$, and $K$) is given, the size could be calculated because the value of $A$ has been identified. On the other hand, these results also show the initial value model proposed in this paper is of vital importance for the size identification.

4.4. Validation of Size Identification Algorithm Based on the Measured Data

The measured field data are the series data of the polished rod position and motor input power in an oil well of Shengli Oilfield of PetroChina. The type of the beam pumping unit is CYJY0-3-53HB with the pumping speed of 2 times/min. The length $N$ of the series is 121. The measured field data are presented in Figure 10. The velocity and acceleration obtained by differentiating the polished rod position are shown in Figure 11. As can be seen from Figure 11, the noises in the velocity and acceleration are loud. For direct illustration of the measurement noise, the polished rod position can be approximated by the truncated
Fourier series [48] when the number of Fourier coefficients is 10. The corresponding approximated polished rod position, velocity, and acceleration are also shown in Figure 11.

![Fourier Series Diagram](image)

**Figure 10.** Measured oilfield data.

![Polished Rod Position, Velocity, and Acceleration Diagram](image)

**Figure 11.** Measured and approximated polished rod position, velocity, and acceleration.

The motor is Y250M-12 with nameplates parameters $P_H = 22$ kW, $n_H = 485$ rpm, $\lambda_k = 2.0$, $\eta_H = 93.0\%$. Since $i_{MB} = 0.037$, the crank velocity can be calculated according to Equations (28) and (33) with the above nameplate parameters. Figure 12 shows the crank angle and acceleration can be obtained by integrating and differentiating the crank velocity.

Now the series data for size identification, $P_R(t)$, $v_R(t)$, $a_R(t)$, $\phi(t)$, $\dot{\phi}(t)$, and $\ddot{\phi}(t)$ have been obtained.
The results indicate that the covariance of the polished rod velocity is appropriate when the identified size error of $R$ is not the smallest although its value is $0.0022$ and $0.0003$, respectively. The reason may be that the noise of the polished rod velocity and acceleration is different from Gaussian noise, and further research is underway.

According to the polished rod position in Figure 10, it has the values of $\lambda_1$ and $\lambda_2$ after solving Equation (36) with the $fminunc$ function of MATLAB. Then, the initial value, lower boundary, and upper boundary can be calculated according to Equations (38) and (40) as follows:

$$
\begin{align*}
    x_0 &= \begin{bmatrix} 0.9375 & 2.2629 & 3.1680 & 3.7786 & 3.3943 \end{bmatrix}^T \\
    x_L &= \begin{bmatrix} 0.6511 & 1.0352 & 2.2000 & 2.5881 & 1.0353 \end{bmatrix}^T \\
    x_U &= \begin{bmatrix} 1.3021 & 4.4001 & 4.4001 & 6.2859 & 8.8001 \end{bmatrix}^T
\end{align*}
$$

According to the differences between the measured and the approximated data shown in Figure 11, the noise covariances of the polished rod position, velocity, and acceleration can be obtained with the $\text{var}$ function of MATLAB, that is, $R_p$ is 0.0003, $R_v$ is 0.0022 and $R_a$ is 0.0233.

When $Q_c = 10^{-7}I_5$, $R_c = R_aI_5$, and $P_0 = 0.15I_5$, the size of the four-bar linkage can be identified according to the Algorithm 1. The results are shown in Table 6. As is shown in Table 6, the maximum error is 7.09% and the minimum error is $-0.15\%$, which indicates the identified results are accurate. To illustrate the effect of the measured covariances, the size is identified when the measured covariance is $R_p$, $R_v$, and $R_a$ respectively. $R_m$ is given by $R_m = \text{diag}(R_p, R_v, R_a, 10^{-3}, 10^{-3})^T$. The results are also presented in Table 6. As can be seen from Table 6, the identified size error of $R_p$ is not the smallest although its value is the smallest, and the error of $R_m$ is the largest because it is a local optimum value.

The results indicate that the covariance of the polished rod velocity is appropriate when the covariances of the polished rod position, velocity, and acceleration are different. The reason may be that the noise of the polished rod velocity and acceleration is different from Gaussian noise, and further research is underway.
To illustrate the effect of the identified results on the polished rod position, Figure 13 shows the measured, real, and identified polished rod positions, in which the real polished rod position is calculated according to the real size values while the identified one is calculated according to the identified size values. As is illustrated in Figure 13, the identified polished rod position is consistent with the real one while the measured polished rod position is different from the real one, which indicates that the algorithm converges to the real polished rod position rather than the measured one.

![Figure 13. Measured, real, and identified polished rod position.](image)

## 5. Results Discussion

### 5.1. Summary of Obtained Results

In this paper, the four-bar linkage size of the beam pumping units is identified based on the simulated data and the measured data. The obtained results are summarized as following:

1. To obtain the crank kinematic data, a speed model of a motor is established according to the Thevenin equivalent circuit of a motor. The validation results indicate that the model is effective. The speed of the motor and the crank kinematic data can be calculated according to the nameplate parameters and the input power of the motor, which is convenient in engineering practice.

2. To obtain the initial value and the boundary conditions of the size identification, an initial value model is established based on the crank slider mechanism and the relative size restrictions of the beam pumping unit. The validation results based on simulated data show that the model is feasible, which is of great importance to the size identification.

3. To avoid the arccosine operation due to the inappropriate initial value in the model, a NCCKF algorithm is proposed. The predicted state among the ordinary CKF algorithm should be checked and then modified by solving the constrained optimization problem. The validation results based on simulated data show that the algorithm is
feasible but may converge to a local optimum value rather than the real value due to
the inappropriate initial value. Therefore, the initial value model is necessary.

(4) To identify the four-bar linkage of a beam pumping unit, a comprehensive size
identification algorithm based on the motor input power and the polished rod position
is proposed. The identified results based on the simulated data show that the identified
sizes can be from the initial values to the corresponding real values within the first
10 s. With the increase in the measurement covariance, the accuracy of identification
decreases. The method to obtain the measured noise covariance of the polished rod
kinematic data is proposed, and the calculated results demonstrate that the noise
covariances of the polished rod position is minimum and that of the acceleration is
maximum. The identified results based on the measured data show that the identified
result is the most accurate only when the noise covariance of polished rod velocity
is utilized.

5.2. Comparison of Different Methods

To better show the advantages of the proposed algorithm, a comparison with several
current identification methods is performed. The identification method is RLS method [49],
EKF [50], UKF [32,51], and RGD method [52], respectively. The needed data are for
Section 4.4 and the measured covariance is $R_n$. In the method of RLS, EKF, and RGD, the
Jacobian matrix is obtained through the complex step differentiation [53]. In RGD method,
the information vector is derived from the Jacobian matrix and $r(0) = 0.01$. In the method
of UKF, $\begin{bmatrix} \alpha & \beta & \kappa \end{bmatrix} = [0.01, 2, 0]$.

The comparison results are shown in Table 7. For comparison purpose, the results of
the proposed algorithm are also listed in Table 7. The mean error in Table 7 denotes the
mean value of the absolute error of different methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>$R$ Error, %</th>
<th>$C$ Error, %</th>
<th>$P$ Error, %</th>
<th>$K$ Error, %</th>
<th>$A$ Error, %</th>
<th>Mean Error, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLS</td>
<td>10.38</td>
<td>-1.40</td>
<td>-4.97</td>
<td>-5.16</td>
<td>-10.78</td>
<td>6.54</td>
</tr>
<tr>
<td>EKF</td>
<td>10.15</td>
<td>-1.36</td>
<td>-5.04</td>
<td>-5.17</td>
<td>-10.76</td>
<td>6.50</td>
</tr>
<tr>
<td>UKF</td>
<td>2.54</td>
<td>-4.10</td>
<td>-10.08</td>
<td>-8.86</td>
<td>-6.72</td>
<td>6.46</td>
</tr>
<tr>
<td>RGD</td>
<td>-4.18</td>
<td>-6.79</td>
<td>-16.72</td>
<td>-14.14</td>
<td>-2.49</td>
<td>8.86</td>
</tr>
<tr>
<td>Proposed</td>
<td>7.09</td>
<td>-0.15</td>
<td>-5.66</td>
<td>-4.62</td>
<td>-6.99</td>
<td>4.90</td>
</tr>
</tbody>
</table>

From the results in Table 7, it can be concluded that the proposed algorithm in the
paper is superior in accuracy to the existing methods.

6. Conclusions

The present paper proposed a size identification algorithm of the four-bar linkage of a
beam pumping unit based on the nonlinear constrained Cubature Kalman filter. According
to the measured motor input power, the crank angle, velocity, and acceleration can be
calculated from the speed model of the motor. The initial value, lower boundary, and upper
boundary can be obtained by the measured polished rod position based on the initial value
model. According to the measured polished rod position, velocity, and acceleration, the
four-bar linkage size is identified by the kinematics model of a beam pumping unit, the state
space model and the Cubature Kalman filter algorithm. The size identification algorithm
is validated based on simulated and measured data, and the identification characteristics
under different measurement covariance are studied. The results show that the established
models and algorithm are feasible. The identified results based on the simulated data show
that the maximum relative error between the identified sizes and the real sizes increases
from 0.23% to 9.37% with the covariance increasing from $10^{-6}$ to $10^{-1}$. The identified results
based on the measured oilfield data indicate that the maximum relative error between
the identified sizes and the actual sizes is 7.09% when the optimal covariance—that is, the measured covariance of the polished rod velocity—is applied. A comparison with several current identification methods indicates the proposed algorithm in the paper is superior in accuracy. Therefore, the designed soft method can be employed in practical applications. It provides essential services for the future intelligent diagnosis and control of the beam pumping units.

Author Contributions: Conceptualization, J.Y.; methodology, J.Y.; software, J.Y.; validation, J.Y.; formal analysis, J.Y.; investigation, J.Y.; resources, D.S.; data curation, D.S.; writing—original draft preparation, J.Y.; writing—review and editing, J.Y.; visualization, J.Y.; supervision, J.Y.; project administration, H.M.; funding acquisition, H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Shandong province Natural Science Foundation of China (Grant No. ZR2021MD067), the Fundamental Research Funds for the Central Universities (Grant No. 22CX03011A).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
2. Song, C.; Liu, S.; Han, G.; Zeng, P.; Yu, H.; Zheng, Q. Edge Intelligence Based Condition Monitoring of Beam Pumping Units under Heavy Noise in the Industrial Internet of Things for Industry 4.0. *IEEE Internet Things* 2022, 1. [CrossRef]
5. Han, Y.; Song, X.; Li, K.; Yan, X. Hybrid modeling for submergence depth of the pumping well using stochastic configuration networks with random sampling. *J. Pet. Sci. Eng.* 2022, 208, 109423. [CrossRef]
16. Ding, F.; Zhang, X.; Xu, L. The innovation algorithms for multivariable state-space models. *Int. J. Adapt. Control* 2019, 33, 1601–1618. [CrossRef]
20. Ye, Y.; Li, Z.; Lin, J.; Wang, X. State-of-charge estimation with adaptive extended Kalman filter and extended stochastic gradient algorithm for lithium-ion batteries. *J. Energy Storage* 2022, 47, 103611. [CrossRef]


22. He, Q.; Stinis, P.; Tartakovsky, A.M. Physics-constrained deep neural network method for estimating parameters in a redox flow battery. *J. Power Sources* 2022, 528, 231147. [CrossRef]

23. Cui, Z.; Wang, L.; Li, Q.; Wang, K. A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network. *Int. J. Energy Res.* 2021, 46, 5423–5440. [CrossRef]


41. Averbukh, M.; Lockshin, E. Estimation of the Equivalent Circuit Parameters of Induction Motors by Laboratory Test. *Machines* 2021, 9, 340. [CrossRef]


